SMBE 2017

Associate Editors Symposium

OS-AE11

Human evolutionary history has increased the role of rare variants in complex phenotypes

Ryan Hernandez [1 2 3,*], Lawrence Uricchio [4], Noah Zaitlen [5]

[1]Bioengineering and Therapeutic Sciences, [2]Quantitative Biosciences Institute, [3]Institute for Human Genetics, UCSF, San Francisco, [4]Biology, Stanford University, Stanford, [5]Lung Biology Center, UCSF, San Francisco, United States

Abstract: Understanding the genetic architecture of complex traits is a central challenge in human genetics. There currently exists a large disparity between heritability estimates from family-based studies and large-scale genome-wide association studies (GWAS), which has been sensationalized as the "missing heritability problem". Among the possible explanations for this disparity are rare variants of large effect that are neither tagged by existing genotyping platforms, nor well imputed from existing reference panels. However, recent population genetic models suggest that the conditions under which rare variants are expected to substantially contribute to heritability may be fairly limited. We have extended existing models of complex traits to incorporate a wider range of plausible evolutionary features, and provide further insights into the role that rare variants play in shaping complex traits. We use these models to investigate the genetic architecture of gene expression levels across European and African individuals using RNA and whole genome sequencing data from the GEUVADIS and 1000 Genomes Projects. In particular, we investigate whether rare variants are likely to be a source of missing heritability in expression across genes. We pioneered a technique for partitioning heritability estimates across allele frequencies using Haseman-Elston (HE) regression. We find that rare variants (MAF < 1%) contribute significantly more heritability than common variants (MAF > 5%) across most genes. This observation suggests that rare variants play a substantial role in the heritability of gene expression patterns, which is inconsistent with neutral evolutionary forces operating on the *cis* regulatory architecture of most genes. We then interrogate multiple large-scale imputed case-control data sets from the to demonstrate that rare variants are also a pervasive factor driving the genetic architecture of several complex diseases. We develop an Approximate Bayesian Computation (ABC) algorithm to infer the evolutionary parameters that can explain these observations, and find a striking relationship between the evolutionary forces that have shaped human genomes and the phenotypic variation we observe.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE2

Through the Looking Glass: Insights into the biodiversity of eukaryotes through phylogenomic analyses and single-cell 'omics'

Laura Katz*

Abstract: We live on a microbial planet as microbes dominate in terms of biodiversity, biomass and biological innovations. Yet many microbial lineages have been understudied, in part because most are currently uncultivable. Our knowledge of microbial diversity is now being transformed by advances in technologies for characterizing genome-scale data. My laboratory has generated genomic and transcriptomic data from diverse microbial eukaryotes, and then combined these data with sequences from public databases to build >13,000 gene trees from ~800 lineages (broadly sampled eukaryotes plus representative bacteria and archaea). We have analyzed the resulting phylogenomic datasets to: 1) create a robust eukaryotic tree of life consisting of five major lineages plus numerous 'orphans'; and 2) develop a hypothesis on the nature of germline and somatic genomes in eukaryotes. Preliminary analyses of single-cell transcriptome and genome data from uncultivable species highlight the power of 'omics' methods for both gene discovery, and for uncovering unusual genome features in lineages sampled across the eukaryotic tree of life. All of this supports Robert W. Hegner's (1938) assertion that: "Alice might have seen something even more wonderful if she had looked through a microscope instead of through a looking glass."

Hegner R.W. 1938. Big Fleas have Little Fleas or Who's Who Among the Protozoa. Baltimore: Williams & Wilkins.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE7

False Negatives Are a Significant Feature of Next Generation Sequencing Callsets

Brenna Henn*

Abstract: Short-read, next-generation sequencing (NGS) is now broadly used to identify rare and common mutations in human population samples and disease cohorts. However, NGS data is known to be error-prone and post-processing pipelines have primarily focused on the removal of spurious mutations or "false positives" for downstream genome datasets. Less attention has been paid to characterizing the fraction of missing mutations or "false negatives" (FN). Here we interrogate several publicly available human NGS autosomal variant datasets using corresponding Sanger sequencing as a truth-set. We examine both low-coverage and high-coverage genomes. We show that the FN rate varies between 3% – 18% and that false-positive rates are considerably lower (<3%) for publicly available human genome callsets like 1000 Genomes. The FN rate is strongly dependent on calling pipeline parameters, not just read coverage. To address this, we design a phylogeny-aware tool [PhyloFaN] which can be used to quantify the FN rate for haploid genomic experiments, without additional generation of validation data. Using PhyloFaN on ultra-high coverage NGS data from both Illumina HiSeq and Complete Genomics platforms derived from the 1000 Genomes Project, we characterize the false negative rate in human mtDNA genomes. The false negative rate for the publicly available mtDNA callsets can be as high as 17 – 20%, even for extremely high coverage haploid data. Our results demonstrate that missing mutations are a significant feature of genomic datasets and imply additional fine-tuning of bioinformatics pipelines is needed. Finally, we consider the downstream implications of a high false negative rate for population genetic analyses, such as inference of admixture from ancient DNA.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE1

Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch

Raúl A. Ortiz-Merino [1], Nurzhan Kuanyshev [2], Stephanie Braun-Galleani [1], Kevin P. Byrne [1], Danilo Porro [2], Paolo Branduardi [2], Kenneth H. Wolfe [1,*]

[1]UCD Conway Institute, University College Dublin, Dublin, Ireland, [2]Dept. of Biotechnology and Biosciences, University of Milano-Bicocca, Milano, Italy

Abstract: It has recently been proposed that the whole-genome duplication (WGD) event that occurred during evolution of an ancestor of the yeast *Saccharomyces cerevisiae* was the result of a hybridization between two parental yeast species that were significantly divergent in DNA sequence, followed by a doubling of the genome content to restore the hybrid's ability to make viable spores. However, the molecular details of how genome doubling could occur in a hybrid were unclear because most known interspecies hybrid yeasts have no sexual cycle. We have discovered that the yeast *Zygosaccharomyces parabailii* provides an almost exact precedent for the steps proposed to have occurred during the *S. cerevisiae* WGD. Two divergent haploid parental species, each with 8 chromosomes, mated to form a hybrid that was initially sterile but regained fertility when one copy of its mating-type locus became damaged by the mating-type switching apparatus. As a result of this damage, the *Z. parabailii* life cycle now consists of a 16-chromosome haploid phase and a transient 32-chromosome diploid phase. Each pair of homeologous genes behaves as two independent Mendelian loci during meiosis.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE6

Comparative genomics of bats: the secret of extended longevity?

Emma Teeling*

Abstract: Of all mammals, bat possess the most unique and peculiar adaptations that render them as excellent models to investigate the mechanisms of extended longevity and potentially halted senescence. Indeed, they are the longest-lived mammals relative to their body size, with the oldest bat caught being 41 years old, living approx. 9.8 times longer than expected. Bats defy the 'rate-of-living' theories that propose a positive correlation between body size and longevity as they use twice the energy as other species of considerable size, but live far longer. The mechanisms that bats use to avoid the negative physiological effects of their heightened metabolism and deal with an increased production of deleterious Reactive Oxygen Species (ROS) is not known, however it is suggested that they either prevent or repair ROS damage. Bats also appear to have resistance to many viral diseases such as rabies, SARS and Ebola and have been shown to be reservoir species for a huge diversity of newly discovered viruses. This suggests that their innate immunity is different to other mammals, perhaps playing a role in their unexpected longevity. Here the potential genomic basis for their rare immunity and exceptional longevity is explored across multiple bat genomes and divergent 'ageing' related markers. A novel blood based population-level transcriptomics approach is developed to explore the molecular changes that occur in an ageing wild population of bats to uncover how bats 'age' so slowly compared with other mammals. This can provide a deeper understanding of the causal mechanisms of ageing, potentially uncovering the key molecular pathways that can be modified to halt, alleviate and perhaps even reverse this process in man.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE14

Bacterial genome evolution: reconciling experiments with phylogeny

Deepa Agashe [1,*]

[1]National Centre for Biological Sciences, Bangalore, India

Abstract: Bacterial genomes show enormous variability in genomic features such as codon bias, tRNA gene pools and genomic GC content. What forces affect their evolution? This has been difficult to answer. Previous work suggests that the answer is strikingly different across timescales. For instance, across bacterial species codon use is strongly correlated with tRNA pools, presumably due to selection for rapid and/or accurate translation. However, in the laboratory, mutants (e.g. with altered codon use) typically adapt via changes in the local sequence context (such as promoter mutations) rather than changes in tRNAs or other aspects of translation. How do we reconcile this gap between micro- and macro- evolutionary patterns, and understand the role of adaptive vs. stochastic processes? To address this, we combine phylogenetic analyses of major bacterial lineages with experimental evolution and whole genome sequencing of mutants with altered GC content, codon use, or tRNA pools. We then integrate our experimental and phylogenetic understanding of the rate, nature and consequences of change in trait values to clarify the evolutionary history of major features of bacterial genomes and predict their future trajectory.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE15

From boom to bust – bacterial adaptation to prolonged survival following resource exhaustion

Sarit Avrani, Evgeni Bolotin, Sophia Katz, Ruth Hershberg*

Abstract: Many bacteria, including the model bacterium *Escherichia coli* can survive for years within spent media, following resource exhaustion. We carried out evolutionary experiments, followed by full genome sequencing of hundreds of evolved clones to study the dynamics by which *E. coli* adapts during the first four months of survival under resource exhaustion. Our results reveal that bacteria evolving under resource exhaustion are subject to intense selection, manifesting in rapid mutation accumulation, enrichment in functional mutation categories and extremely convergent adaptation. Our results further demonstrate that such adaptation is not limited by mutational input. Indeed, mutational input appears to be high enough to enable bacteria to rapidly adapt, in a highly convergent manner and with great temporal precision through fluctuations in allele frequencies. Finally, we demonstrate that due to antagonistic pleiotropy and mutation accumulation, survival under resource exhaustion can severely reduce a bacterium's ability to grow exponentially, once resources are again available. Combined, our results shed light on bacterial adaptation to long-periods of resource exhaustion and on the consequences such adaptation has on the genetic makeup of individual bacteria and on patterns of genetic variation within bacterial populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

POA-7

Unusual patterns of incomplete sweeps suggest frequency-dependent positive selection in African populations of Drosophila melanogaster

Yuseob Kim [1,*], Ha My Vy [1]

[1]Ewha Womans University, Seoul, Korea, Republic Of

Abstract: While the great impact of positive selection on the genomic level and pattern of sequence polymorphism is well confirmed in many species, it still remains a challenge to elucidate how beneficial mutations arise and propagate in a population and how selective pressures on mutant alleles are structured over space and time. By identifying "sweeping haplotypes (SHs)" that are increasing (or have increased) rapidly in frequency, and surveying the geographic distribution of SH frequencies, we can infer how selective sweeps unfold in time and thus which modes of positive selection underlie those sweeps. Using the population genomic data of African *Drosophila melanogaster* provided by DPGP project, we identified SHs from 39 candidate loci under selection. At more than seven loci, SH frequencies are similar across multiple populations, which cannot be explained unless some mechanism of frequency-dependent positive selection, such as heterozygote advantage, is invoked considering the reasonable range of migration rates between African populations. We also identify several loci under soft selective sweeps. In one locus, which includes CG30007 gene that exhibits a high level of nonsynonymous polymorphism, many independent SHs are found over multiple populations but always together with the ancestral haplotype block. This complex pattern is compatible with a large number of mutational targets in a gene and the frequency-dependence of new variants. We will discuss biological models that predict such frequency-dependent behavior of new beneficial mutations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

POA-6

The complex symbiome of the carmine cochineal Dactylopius coccus

Arturo Vera-Ponce de León [1], Ernesto Ormeño Orrillo [2], Shamayim Tabita Ramirez-Puebla [1], Rafael Bustamante [1], Mónica Rosenblueth [1], Julio Martinez Romero [1], Esperanza Martinez-Romero [1,*]

[1]Ecological Genomics, Genomic Science Center, UNAM, Cuernavaca, Mexico, [2]Universidad La Molina, Lima, Peru

Abstract: *Dactylopius coccus* is a scale insect that feeds on cactus sap. This insect is the main source of carmine dye used in textile, pharmaceutical and food industries. Bacterial endosymbionts of *D. coccus* comprise two *Wolbachia* strains and a β-proteobacteria named as *Candidatus* Dactylopiibacterium carminicum (*D. carminicum* from here on), present in all the five *Dactylopius* species sampled. In addition there are diverse fungi (Vera Ponce de Leon et al 2015) and spiroplasma. The aim of our work was to study *D. carminicum* microbiome. As we are unable to grow *D. carminicum* or spiroplasma in laboratory culture media, their genomes were assembled from three metagenomes obtained from *D. coccus*. A 3.6 Mb whole genome reconstruction of *D. carminicum* was possible from the metagenomic data. The spiroplasma genome is nearly 1 MB and a phylogenomic analysis showed that the cochineal symbiont may correspond to a new species of spiroplasma. RT-PCR revealed the expression of *D. carminicum nifH* genes in hemolymph, ovaries and eggs of the cochineal. Moreover, acetylene reduction assays showed nitrogen fixing activity in the same tissues. Fluorescent *in situ* hybridization using *D. carminicum* 16S rRNA probes localized this bacterium in embryo and ovaries. Our results indicated that *D. carminicum* harbors a complex symbiome with maternally inherited members that seemingly fulfill different functions in the host.

Vera-Ponce de León A, Sanchez-Flores A, Rosenblueth M, Martínez-Romero E. Fungal community associated with Dactylopius (Hemiptera: Coccoidea: Dactylopiidae) and its role in uric acid metabolism. Front Microbiol. 2016 7:954. doi: 10.3389/fmicb.2016.00954.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE4

Sequence amplification via cell passaging creates spurious signals of positive adaptation in influenza virus H3N2 hemagglutinin.

C D. McWhite [1], A G. Meyer [2], C O. Wilke [2,*]

[1]Department of Molecular Biosciences, [2]Department of Integrative Biology, The University of Texas, Austin, United States

Abstract: Clinical influenza A virus isolates are frequently not sequenced directly. Instead, a majority of these isolates (~70% in 2015) are first subjected to passaging for amplification, most commonly in non-human cell culture. Here, we find that this passaging leaves distinct signals of adaptation, which can confound evolutionary analyses of the viral sequences. We find distinct patterns of adaptation to Madin–Darby (MDCK) and monkey cell culture absent from unpassaged hemagglutinin sequences. These patterns also dominate pooled datasets not separated by passaging type, and they increase in proportion to the number of passages performed. By contrast, MDCK–SIAT1 passaged sequences seem mostly (but not entirely) free of passaging adaptations. Contrary to previous studies, we find that using only internal branches of influenza virus phylogenetic trees is insufficient to correct for passaging artifacts. These artifacts can only be safely avoided by excluding passaged sequences entirely from subsequent analysis. We conclude that future influenza virus evolutionary analyses should appropriately control for potentially confounding effects of passaging adaptations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE5

Combinatorial mutagenesis of positively selected residues reveals negative epistasis and super-restrictor phenotypes in an antiviral protein

Rossana Colon-Thillet [1], Maria Gutierrez [2], Richard McLaughlin [3], Michael Emerman [1], Harmit Malik [4,*]

[1]Basic Sciences & Human Biology, Fred Hutchinson Cancer Research Center, Seattle, [2]Biology, University of Texas El Paso, El Paso, [3]Basic Sciences, [4]Basic Sciences & HHMI, Fred Hutchinson Cancer Research Center, Seattle, United States

Abstract: The innate arm of mammalian immunity encodes hundreds of antiviral proteins that act cell-autonomously to block viral replication, often by binding virally encoded proteins. To respond to rapid viral evolution, antiviral genes evolve at an accelerated rate. Signatures of diversifying selection can be used to successfully predict antiviral protein surfaces that are used to recognize viral pathogens. Functional studies have demonstrated that the binding specificity to viral targets is governed by only a few rapidly evolving residues at the interaction interfaces. For instance, we previously showed that single amino acid residue changes in positively selected residues of Loop L4 dictate the antiviral specificity of the broadly antiviral protein, MxA. Using a combinatorial mutagenesis screen, we built a library of human MxA variants that encodes every possible five amino acid combination at the rapidly evolving sites in L4, and assessed their antiviral functionality against the Thogoto virus (THOV), a tick-borne rodent orthomyxovirus. Our screen revealed that all active MxA variants (~5%) had a strict preference for phenylalanine (F), tryptophan (W) or tyrosine (Y) at position 561. However, consistent with negative epistasis, a number of MxA variants recovered were inactive despite possessing a F, Y, W at residue 561. Finally, we found a rare set of variants with enhanced antiviral activity against THOV when compared to human MxA. Use of combinatorial mutagenesis in residues subject to diversifying selection during evolution could provide a means to potentially increase antiviral efficacy of proteins like MxA, for therapeutic use.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE12

TimeTree: A resource for timelines, timetrees, and divergence times

Stephen Hedges*, Glen Stecher [1], Michael Suleski [1], Sudhir Kumar [1]

[1]Temple University, Philadelphia, United States

Abstract: Evolutionary information on species divergence times is fundamental to studies of biodiversity, development, and disease. Molecular dating has enhanced our understanding of the temporal patterns of species divergences over the last five decades, and the number of studies is increasing quickly due to an exponential growth in the available collection of molecular sequences from diverse species and large number of genes. Our *TimeTree* resource is a public knowledge-base with the primary focus to make available all species divergence times derived using molecular sequence data to scientists, educators, and the general public in a consistent and accessible format. Here, we report a major expansion of the *TimeTree* resource, which more than triples the number of species (>97,000) and more than triples the number of studies assembled (>3,000). Furthermore, scientists can access not only the divergence time between two species or higher taxa, but also a timetree of a group of species and a timeline that traces a species' evolution through time. The new timetree and timeline visualizations are integrated with display of events in earth and environmental history over geological time, which will lead to broader and better understanding of the interplay of the change in the biosphere with the diversity of species on Earth. The next generation *TimeTree* resource is publicly available online at http://www.timetree.org.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

POA-4

Nucleotide vs clonal coalescence in bacteria

Daniel Falush*

Abstract: Effective population size is one of the most important concepts in population genetics but also amongst the most enigmatic, particularly for microorganisms. For bacteria, effective population size can be measured for the organism, i.e. based on the rate of coalescence of clonal lineages, or for the genome, i.e. based on the average rate of coalescence of DNA sequences. Under neutral models the two approaches should give equivalent answers. I describe the assumptions required to measure and compare these rates. There are >50fold differences in the value of the two measurements in Vibrio parahaemolyticus, with smaller but still substantial differences in other species. The differences in estimates should be very informative about how natural selection acts to structure bacterial populations but several assumptions need to be more fully tested in order to narrow down the likely causes. I discuss initial attempts to test these assumptions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

POA-3

Support for lungfish as the closest relative of tetrapods by using slowly evolving ray-finned fish as the outgroup

Naoko Takezaki*, Hidenori Nishihara [1]

[1]Department of Life Science and Technology, Tokyo Institute of Technology, Yokohama, Japan

Abstract: In a previous analysis of the phylogenetic relationships of coelacanths, lungfishes and tetrapods, using cartilaginous fish as the outgroup, the sister relationship of lungfishes and tetrapods was constructed with high statistical support. However, using as the outgroup ray-finned fish, which are more taxonomically closely related to the three lineages than cartilaginous fish, the sister relationship of coelacanths and tetrapods was most often constructed depending on the methods and the data sets, but the statistical support was generally low except in the cases in which the data set including a small number of species was analyzed. In this study, instead of the fast evolving ray-finned fish, teleost fish, in the previous data sets, by using two slowly evolving ray-finned fish, gar and bowfin, as the outgroup, we showed that the sister relationship of lungfishes and tetrapods was reconstructed with high statistical support. In our analysis the evolutionary rates of gar and bowfin were similar to each other and one third to one half of teleost fish. The difference of the amino acid frequencies of the two species with other lineages were larger than those of teleost fish. This study provides a strong support for lungfishes as the closest relative of tetrapods and indicates the importance of using an appropriate outgroup with small divergence in phylogenetic construction.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

POA-10

The complex evolution of Myxozoa (Cnidarian) mitochondrial genomes

Dorothee Huchon [1,*], Dayana Yahalomi [1], Arik Diamant [2], Jerri Bartholomew [3], Stephen Atkinson [3], Paulyn Cartwright [4]

[1]Zoology, Tel-Aviv University, Tel-Aviv, [2]Israel Oceanographic and Limnological Research, Eilat, Israel, [3]Department of Microbiology, Oregon State University, Corvallis, [4]Department of Ecology & Evolutionary Biology, The University of Kansas, Lawrence, United States

Abstract: Myxozoans are microscopic cnidarian parasites whose infections cause substantial damage to fish aquaculture. They have a very simple organization and a complex life cycle that typically alternates between two hosts: a fish and an annelid. Myxozoa is comprised of thousands of species and currently, little is known about their genome evolution. To gain further insight into their evolution, we sequenced the mitochondrial (mt) genome of representative species representative of the myxozoan diversity and their sister clade *Polypodium hydriforme*.

Our preliminary results suggest that the ancestral myxozoan mt genome consisted of a single circular molecule, which fragmented into 2 to 8 circular chromosomes in extant lineages. Within a species, the different chromosomes share almost identical non-coding region of length up to ~15 kb. Consequently, Myxozoa includes the largest described animal mt genomes. The high homology between the shared non-coding regions is likely to be maintained by gene conversion, which is supported by their high GC content. Additionally, we could identify in some species the presence of chimeric mitochondrial chromosomes, suggesting that the homologous non-coding regions recombine.

In all Myxozoa species studied the protein coding genes show an unusually high rate of sequence evolution and possess little similarity to their cnidarian homologs. Only five protein coding genes could be identified. Remarkably, our analyses suggest the absence of tRNA genes within the mt chromosomes of Myxozoa. This observation is supported by the absence in the nuclear genome of key proteins involved in the mitochondrial translational machinery such as the aminoacyl-tRNA synthetase genes or the mtRNaseP subunit MRRP3 gene. These observations confirm the remarkable plasticity of myxozoan mt genomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE3

Compositional heterogeneity, Substitutional saturation and the problem of early animal evolution

Davide Pisani*

Abstract: The relationships at the root of the animal tree of life have proven hard to resolve, with the current debate focusing on whether sponges (phylum Porifera) or comb jellies (phylum Ctenophora) represent the sister group of all other animals. Modelling of the amino acid substitution process is at the core of this debate because previous studies showed that Porifera tends to emerge as sister to all other animals (Porifera-sister), when site-specific amino acid differences are modelled, while Ctenophora emerges as the sister group of all the other animals (Ctenophora-sister) when site-specific amino acid differences are not modelled.
Results will be presented showing that models routinely used in early animal phylogenetics fail to adequately describe site- and lineage-specific compositional heterogeneity across key datasets. Data recoding can be used to concomitantly reduce compositional heterogeneity and saturation, and recoded datasets, when modelled using the most adequate substitution model, invariably provide outgroup independent support for Porifera as the sister group of all the other animals.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE8

Inferring rates and length-distributions of indels using approximate Bayesian computation

Tal Pupko* and Eli Levy Karin, Haim Ashkenazy, Reed A. Cartwright

Abstract: The most common evolutionary events at the molecular level are single-base substitutions, and insertions and deletions of short DNA segments (indels). A large body of research has been devoted to develop probabilistic substitution models and to infer their parameters using likelihood and Bayesian approaches. In contrast, relatively little has been done to model indel dynamics, probably due to the difficulty in writing explicit likelihood functions. Here, we contribute to the effort of modeling indel dynamics by presenting SpartaABC, an approximate Bayesian computation (ABC) approach to infer indel parameters from sequence data (either aligned or unaligned). SpartaABC circumvents the need to use an explicit likelihood function by extracting summary statistics from simulated sequences. First, summary statistics are extracted from the input sequence data. Second, SpartaABC samples indel parameters from a prior distribution and use them to simulate sequences. Third, it computes summary statistics from the simulated sets of sequences. By computing a distance between the summary statistics extracted from the input and each simulation, SpartaABC can provide an approximation to the posterior distribution of indel parameters as well as point estimates. We study the performance of our methodology and show that it provides accurate estimates of indel parameters in simulations. We next demonstrate the utility of SpartaABC by studying the impact of alignment errors on the inference of positive selection. A C++ program implementing SpartaABC is freely available in spartaabc.tau.ac.il.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

POA-2

Learning site-specific biochemical property preferences and constraints from comparative sequence analysis.

Sergei Pond [1,*], Stephanie Spielman [1], Joel Wertheim [2], Konrad Scheffler, Steven Weaver [1], Ben Murrell [2]

[1]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, [2]Medicine, University of California San Diego, San Diego, United States

Abstract: Probabilistic models of protein evolution are used to constructsubstitution matrices for protein sequence alignment, to inferphylogenetic trees, and to infer which sites in genomes have beentargeted by natural selection.
While existing models have become adept at answering the questions o where in the sequence, or when in the evolutionary history, selection may have acted, they are surprisingly limited in answering th question of how this action was realized. At a particular site, sequence evolution conserves biochemical properties which must be preserved for functional reasons, and changes the properties when doing so confers a selective advantage. This isintuitively apparent, for example because only a few amino acids are ever observed at a given site in most protein-coding genes.
Our property informed model of evolution (PRIME) encodes this intuition in a formalphylogenetic maximum likelihood hypothesis testing framework, and determine the importance of various physico-chemical properties such as polarity, hydropathy, and volume, recognizing that this importance varies from site to site.We demonstrate that previous attempts to incorporate physico-chemical properties into sequence evolution have not been very successful, largely because they did not allow property importance to vary from one site to another, whereas structural biology makes it abundantly clear that it does. We also examine how the choice of biochemical properties to model affects inference, whether or not there may be a "universal" set of properties suitable for evolutionary inference. PRIME is a natural extension of existing methods for studying adaptive evolution, because it deconstructs the composite action of natural selection into property-based components, and it does so without the use of any stuctural information. PRIME is also a logical complement to experimentally informed models of evolution, e.g., deep mutational scanning, since it extracts information about long-term evolutionary forces, as opposed to finely characterizing the accessible sequence space around the current state of a protein.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

POA-1

Tempo and Mode of Mutation Rate Evolution in C. elegans Under 400 Generations of Minimal Selection

Charles Baer [1,*], Ayush Saxena [1], Matthew Salomon [1], Chikako Matsuba [1], Shu-Dan Yeh [1]

[1]Department of Biology, University of Florida, Gainesville, United States

Abstract: The rate and molecular spectrum of mutations varies within and among genomes, species and higher taxa. The understanding of the mechanistic causes of variation is far from complete; the understanding of evolutionary causes is even more rudimentary. It has been posited that individuals in poor physiological condition will experience higher rates of mutation than will individuals in good condition. Since deleterious mutations lead to poor condition, the possibility exists that individuals carrying a high load of deleterious mutations will experience an elevated mutation rate. We tested that hypothesis with a set of "second-order mutation accumulation" (o2MA) lines of the nematode *Caenorhabditis elegans*. MA lines that had accumulated mutations under minimal selection for ~250 generations ("first-order MA lines", o1MA) were sorted into high-fitness and low-fitness groups, replicated into new sets of o2MA lines, and allowed to accumulate mutations for another ~150 generations of minimal selection. Whole-genome sequencing of 48 o2MA lines and their o1MA ancestors did not detect a significant effect of initial fitness on the subsequent base-substitution rate. However, there was significant variation in second-order mutation rate among o1MA lines. The deletion rate of low-fitness o2MA lines was less than that of high-fitness o2 lines, but that result is plausibly due to synergistic epistasis rather than different mutational processes. Multiple logistic regression of mutability on a set of predictor variables revealed that local three-base nucleotide context is the most important predictor of mutability, but that GC content of the 1 Kb surrounding a site and – importantly – local recombination rate are also significant predictors. Mutability explains a large fraction of the variance in standing nucleotide diversity. Second-order mutation rate was slightly but significantly greater than the first-order rate, consistent with the "drift-barrier" hypothesis of mutation rate evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE10

Adaptive remodeling of the transcriptional response to drought stress in Arabidopsis lyrata

Juliette de Meaux* on behalf of Plant Molecular Ecology, University of Cologne, Germany, Fei He on behalf of Plant Molecular Ecology, University of Cologne, Germany, Agustin Arce on behalf of Plant Molecular Ecology, University of Cologne, Germany, Gregor Schmitz on behalf of Plant Molecular Ecology, University of Cologne, Germany, Andreas Beyer [1] and Plant Molecular Ecology, University of Cologne, Germany

[1]CECAD, University of Cologne, Cologne, Germany

Abstract: To describe how stress responses evolve, we undertook a time serie analysis of plant transcriptome responses to progressive dehydration in the species *Arabidopsis lyrata*, which is robust to drought stress, and compared it to the response displayed by the more sensitive sister species *A. halleri*. Using the model plant species A. thaliana as an outgroup, we show that while *A. halleri* displays a response that converges towards an ancestral response, *A. lyrata* has evolved a novel transcription profile. Trans-acting mutations predominate at early steps of the response, with cis-acting mutations gaining in importance at later stages oft he response. To investigate the footprint of polygenic selection on the timing and dynamics of the expression response, we analyzed the distribution of cis-acting mutations derived in each lineage, an approach that we had previously used to demonstrate selection on heavy metal genes in *A. halleri*. We show that the response displayed by *A. lyrata* 6h after initiation of the stress carries the signature of polygenic selection on the GO category …regulation of response to stress". We contrast our results with a population genomics survey of selection coefficients throughout the genome. This work pioneers a novel and and innovative approach to identify the action of natural selection on the polygenic underpinnings of a complex adaptive trait.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

POA-11

Modelling PCR stochasticity and its effect on quantitative NGS experiments

Arndt Von Haeseler*, Florian Pflug [1]

[1]CIBIV, Vienna, Austria

Abstract: Many protocols in modern-day biology use next-generation sequencing (NGS) as a quantitative method, i.e. to measure the abundance of particular DNA molecules. Then, any molecule that remains unsequenced causes a measurement error, and if this affects molecules non-uniformly, results are systematically biased. A major source of such biases is the Polymerase Chain Reaction (PCR), used to amplify DNA prior to sequencing. If it can be adequately modelled, its biases can be predicted and corrected for. Different models of PCR haven been proposed, but none have yet found their way into standard analysis pipelines, owing to a lack of parameter estimates for specific conditions. We thus focus on describing a model whose parameters can be estimated from actual experimental data, while still capturing the main source of biases. We show that this is achieved by viewing PCR as a branching process which, during each cycle, duplicates each DNA molecule with a certain probability, called the reactions efficiency. We combine this model with a simple model of the sampling behaviour of NGS and apply it to published RNA-Seq data. We demonstrate that the reaction efficiency can be estimated from the data, and that the data matches the models predictions well. In particular, we find that the model explains the main observed stochastic effects. Finally, we explore how well we can correct for unobserved molecules, and how much this improves the accuracy of the measured gene transcript abundances.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE13

A Public Goods Perspective On Evolution

James Mcinerney*

Abstract: The "standard model" for thinking about evolution is the phylogenetic tree. There is no question of the usefulness and appropriateness of phylogenetic trees in many situations, however recombination has played a large part in structuring genes and genomes and this net-like history is more appropriately analysed using network models. Several network approaches have bene used in order to understand recombination and in this talk I will outline their relationships to one another and what can be achieved by each approach. I will then outline how sequence similarity networks and bipartite graphs can be used to explore the evolutionary history of highly recombinogenic molecules. Overall, we find that DNA sequences can act as *bona fide* public goods, which means they are usually not excludable or rivalrous and there is little additional cost of production in order to make them publicly available. This "goods-thinking" model is a useful additional metaphor for evolutionary biology.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE9

Characteristics of evolutionarily conserved noncoding regions in vertebrate genomes

Naruya Saitou*

Abstract: A considerable region of eukaryote genomes is noncoding. Majority of them are junk DNA and do not have functions. If we find evolutionary conservation, however, these conserved regions are expected to have some function which is protected through purifying selection. From the initial stage of molecular evolutionary studies, noncoding regions were suspected to be involved in gene regulation. Now it is becoming clear that at least some noncoding regions play important roles in gene regulation. Therefore, conserved noncoding sequences (CNSs) are likely to be important from the functional point of view. CNS analyses have been proved to be powerful for detecting regulatory elements. We compared genome sequences of vertebrates, various mammalian orders, especially primates for searching CNSs and their characteristics (Takahashi & Saitou 2012; Matsunami & Saitou 2013; Babarinde & Saitou 2013, 2016; Hettiarachchi et al. 2013, 2016; Saber et al. 2016). One common feature emerged from analyses of these CNSs is that they are often closely located to transcription factor genes and their GC contents are different from genomic averages. Interestingly, vertebrate CNSs have lower GC content than their genomic average, while typical eukaryote CNSs have higher GC content. Amniote-specific CNSs were found to keep physical distance with the nearby coding gene during their evolution to mouse and human. We recently analyzed hominoid- and hominid specific CNSs, and found that some of these CNSs emerged through positive selection. These features indicate the importance of CNSs on phenotypic evolution of vertebrates.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

OS-AE16

Combining disease data and molecular evolution to better understand both

Alex Griffing, Eric Stone [1], Jeffrey Thorne [2,*]

[1]ANU-CSIRO Centre for Genomics, Metabolomics, and Bioinformatics, Australian National University, Canberra, Australia, [2]Biological Sciences and Statistics, North Carolina State University, Raleigh, United States

Abstract: Inviable genotypes represent "holes" in the adaptive landscape. Because the history of sequences that result in observed genotypes must constitute a path that avoids fitness holes, knowledge of the hole locations can improve evolutionary inferences. Molecular evolution on a holey adaptive landscape can be considered from the perspective of a particular codon location. At one point in time, certain amino acids encoded at this location might yield fitnesses of zero. Later, epistatic changes elsewhere in the genome might eliminate the holes. Because an amino acid at some protein site might be inviable in some environments and viable in others, it is also possible that a fitness hole disappears or appears if the environment changes.
We developed a model-based approach to analyze aligned interspecific protein-coding sequences. It uses disease-mutation data to provide information regarding where fitnesses holes are. It uses observed (wildtype) gene sequences to provide information about where fitness holes are not. Our evolutionary model has two layers. One describes the locations of holes in the adaptive landscape and how the locations change over time. The other describes the state of the protein-coding sequence and how it traverses the adaptive landscape. Inferences from the approach reflect the fact that ancestral lineages have avoided fitness holes. In addition, the approach can be employed to estimate rates at which fitness holes appear and disappear. We have been applying our approach to the evolution of mammalian TP53 genes. We will discuss our approach, our inferences, and our plans.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

POA-8

Evolution of Regulation in Cell to Cell Communication

Janine Quijano, Robert Wisotzkey, Stuart Newfeld [1,*]

[1]School of Life Sciences, Arizona State University, Tempe, United States

Abstract: Cell to cell communication is a fact of life for multicellular organisms. It is required during development to insure the coordinated differentiation of progenitor cells into species-specific patterns of organs and tissues. It is required for maintaining homeostasis in adults and when disrupted can lead to problems such as loss of mitotic control and cell death. In humans these problems manifest themselves as cancer and neurodegenerative disease. Thus, the importance of understanding how cell to cell communication is regulated. In a testament to the dexterity of natural selection, there are only a handful of cell to cell communication pathways. These govern countless developmental and homeostatic events in the lifetime of a single multicellular organism and the collective lifetimes of all multicellular organisms. One of these is the Transforming Growth Factor-beta pathway (TGF-beta). Highly conserved since the dawn of multicellularity, the pathway directs developmental events in all metazoan species and in humans its signal transducing Smad proteins act as tumor suppressors. Given the billion year conservation of the TGF-beta pathway, how is it that species-specificity is imposed? Utilizing molecular genetics in Drosophila together with bioinformatics, two examples of differential TGF-beta pathway regulation between flies and mammals have been discovered. These studies of epigenetic proteins and ubiquitin ligases identified two evolutionarily new genes that have been adopted by TGF-beta developmental networks that may also function as tumor suppressors. Overall, these examples reveal that the species-specific regulation of cell to cell communication is a fundamental feature of organismal diversity and human disease.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Associate Editors Symposium

POA-9

Theoretical foundation of the RelTime method for estimating divergence times

Koichiro Tamura [1,*], Sudhir Kumar [2]

[1]Research Center for Genomics and Bioinformatics, Tokyo Metropolitan University, Hachioji, Japan, [2]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, United States

Abstract: The RelTime approach was originally presented as an algorithm for generating an ultrametric tree from a molecular phylogenetic tree with branch lengths, where evolutionary rate varies from branch to branch. The RelTime algorithm has been found to produce accurate estimates of divergence times in computer simulations and empirical data analyses, and its performance compares favorably with computation-intensive Bayesian approaches. Here, we present theoretical underpinnings of the RelTime approach and show that it employs the principle of minimum net rate change needed to explain the observed branch lengths in a phylogeny. This information paves the way for broader use of RelTime approach for estimating evolutionary rates and divergence times not only for molecular sequence alignments, but also for other types of molecular data as well as non-molecular data.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

POB-62

Molecular timetrees of Chelicerata recover monophyly of Arachnida and suggests an early colonization of land

Jesus Lozano-Fernandez [1,*], Alastair R. Tanner [2], Gregory Edgecombe [3], Davide Pisani [1]

[1]School of Earth and Biological Sciences, [2]School of Biological Sciences, University of Bristol, Bristol, [3]Department of Earth Sciences, Natural History Museum, London, United Kingdom

Abstract:

Animal life have marine origins, with only few phyla completing their entire life cycle outside water. The process through which organisms adapt to life on land is known as terrestrialization and it is one of the most extreme cases of adaptation to a new habitat that happened in animal history. The chelicerates (pycnogonids, horseshoe crabs, spiders, scorpions) are an ancient group of arthropods, with an astonishing fossil record dating back to Cambrian, and includes the second largest clade of fully terrestrial organisms, the arachnids. Morphological phylogenies support a single land colonization event by placing marine horseshoe crabs as sister group of arachnids, but phylogenomic studies nest this aquatic lineage within Arachnida (implying multiple terrestrialization events). To identify how many times and when arachnids adapted to life on land we need to assess chelicerate phylogeny, and its evolutionary timescale.

Here we present a timescale of chelicerate evolution based on an expanded multigene dataset that covers most chelicerate diversity (>240 genes and >100 taxa) and using the largest set of fossil calibrations to date. Our results recover monophyly of Chelicerata, Euchelicerata and Arachnida. Aracahnida is conformed by two megaclades, one composed by Pseudoscorpions and Arachnopulmonates (scorpiones + spiders and closely related allies). The second megaclade contains a poorly supported association between Opiliones and Ricinulei, allied with a clade formed by monophyletic Acari sister to Solifugae. Our results reconcile previous results based on morphology and molecular evidence, and suggest a Cambrian-Ordovician colonization of land by Arachnids, substantially predating trace or body fossil evidence.

Expanded summary*:

Molecular palaeobiology and comparative genomics of arthropod terrestrialization

The ancestral habitat where animals emerged is marine and only four phyla including lineages that can complete every phase of their life cycle outside of water–saturated environments (from moisture films to the oceans). These phyla are the Vertebrata (with the reptiles, birds and mammals), the Mollusca (with the land snails and the slugs), Onychophora (velvet worms) and the Arthropoda (e.g. insects, spiders, scorpions, centipedes). The process through which animals adapted to life on land is referred to as terrestrialization and it is one of the most fascinating unresolved problems in evolutionary biology. The crossing of the water–land barrier was the most extreme case of adaptation to a new environment in animal history. In fact, the difference between the sea and the subaereal environment is so extreme that astrobiology uses terrestrialization as an analog to study how life could adapt to a new planet.

Arthropods appeared in the Cambrian ~520 million years ago and include three subphyla: Chelicerata (e.g. spiders), Myriapoda (e.g. centipedes) and Pancrustacea (e.g. insects and crustaceans). The oldest subaerial arthropod traces are from the late Cambrian (~500 Ma old) and by the late Silurian (~428 Ma ago) all subphyla included terrestrial lineages. Arthropods represent the largest majority (~80%) of biodiversity on our planet, and the largest majority of arthropods are terrestrial. This illustrates how important the process of adaptation to life on land was in animal history and as a generator of biodiversity. In Arthropoda, there have been a minimum of three ancient (Palaeozoic) terrestrialization events: that of the Hexapoda, that of the Myriapoda and that of the Arachnida. The fact that they colonised the land multiple times independently allow rigorous comparison of the alternative solutions adopted by the different groups to the same adaptive challenge.

The relationships among the arthropod subphyla has long been debated, and current consensus suggests that Myriapoda and Pancrustacea form a clade (Mandibulata), with Chelicerata representing their sister taxon. However, relationships *within* the subphyla

are still debated, and we still do not know how arthropods adapted to life on land. In my current research I try to elucidate the evolutionary history of arthropods and their subphyla by implementing a molecular palaeobiological approach, where molecular data from living organisms is combined with fossil data to understand changes through time within the context of deep time. The goal of my project is to combine molecular and morphological data, infer a complete timetree for the arthropods (including both extant and fossil lineages) and use it to test hypotheses of their evolution. My investigations try to answer 'how' and 'when' these organisms have colonized the land by means of inferring the arthropod relationships and identifying genomic adaptations involved in this terrestrialization process. So far I have been studying in depth the origin of Hexapoda and Arachnida.

Chelicerates (e.g. spiders, scorpions, mites) represent an ideal model system to investigate the *tempo* and *mode* of early animal terrestrialization, because they were the first animals to become abundant in the terrestrial fossil record. I am particularly interested in the comparative genomics of this group because they contain marine and land representatives, which allows investigating genes that has been evolutionary selected in terrestrial lineages. I am also interested in the origin of scorpions, as it is an ancient terrestrial lineage with an extensive Palaeozoic fossil record (~ 435 Mya) with an uncertain affinity to the rest of Arachnids, in which we still do not know how they colonised the land.

My research deals with ancient colonizations of land environment in different lineages that are phylogenetically related, which provide a unique opportunity to study in parallel this extreme adaptation process. The results of this research are significant for establishing the evolutionary history of the most diverse living group as well as for understanding the biology and mechanisms of this adaptive process. The knowledge generated from study this of habitat colonization process could be helpful in understanding the biology of invasive species, or the mechanisms involved in adaptation to a new environment. Moreover, chelicerates include pests (e.g. the spider mites) and species of biomedical relevance (e.g. the ticks – vectors of lyme disease). By identifying chelicerate–specific genomic adaptation to life on land, this project could identify potential chelicerate–specific drug targets which may help the development of lineage–specific pesticides with low incidence on economically important arthropods, like the declining bees.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

OW-CH1

Calibrating the microbial Tree of Life

Lance Jones [1], Yagya Sharma [1], Brandon Hanna [1], Alex L'Esperance [1], Mandarke Pai [1], Fabia Ursula Battistuzzi*

[1]Oakland University, Rochester, United States

Abstract: The process of estimating the timing of the tree of life faces many challenges that range from phylogenetic uncertainties to methodological issues. The microbial part of the tree of life is particularly affected by these issues because of their complex evolutionary histories, that are difficult to accurately model (e.g., horizontal gene transfer), and their sparse fossil record that severely limits the number of points to use as anchors for molecular clock estimates (times and rates). However, unicellular eukaryotes and prokaryotes represent the majority of lineages in the Tree of Life and, therefore, a timetree of life cannot be achieved without solving the puzzle of their evolutionary history. Two of the biggest challenges to microbial molecular clocks are related to the boundaries and numbers of available calibrations. Few nodes in microbial phylogenies can be (more or less) confidently timed from non-molecular data (e.g., biomarkers) and, of these, most have large boundary ranges or a single boundary (either a minimum or a maximum-only). This scenario is known from simulation and empirical data to be susceptible to artifacts and, therefore, requires accurate evaluation to determine the validity of its outcomes. Here, we address this issue using the latest development in molecular clock methods to evaluate the strength and weaknesses of calibration points commonly used within the microbial Tree of Life.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

OW-CH2

Hierarchical Bayesian models for integrating paleontological and neontological data

Tracy Heath [1,*]

[1]Ecology, Evolution, & Organismal Biology, Iowa State University, Ames, United States

Abstract:

Understanding macroevolutionary processes and evolution in deep time requires data from the fossil record. In recent years, advances in phylogenetic inference methods have provided ways to integrate fossil and extant taxa. These approaches allow simultaneous estimation of the divergence times and phylogenetic relationships of extant and fossil species, thus making full use of morphological and temporal data, rather than just molecular sequence data from living species. In particular, diversification models that accommodate fossil sampling incorporate more data from the fossil record than traditional node-calibration approaches. As a result, these approaches provide robust estimates of node ages and better representations of statistical uncertainty. I will highlight our recent and ongoing work using hierarchical Bayesian models to estimate species phylogenies and divergence times. Both simulation and empirical studies demonstrate how making full use of available fossil data and properly modeling lineage sampling and diversification improve estimates of species divergence times.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

POB-58

A coarse-graining, ultrametric approach to resolve the phylogeny of prokaryotic strains with frequent recombination

Tin Yau Pang*

Abstract:

Homologous recombination happens when a foreign DNA stretch replaces a similar stretch on the genome of a prokaryotic cell. For a genome pair, recombination affects their phylogenetic reconstruction in multiple ways: (i) a genome can recombine with a DNA stretch that is similar to the other genome of the pair, thereby reducing their pairwise sequence divergence; (ii) a genome can also recombine with a stretch from an outgroup-genome and increase the pairwise divergence. Most phylogenetic algorithms cannot account for recombination; while some do, they cannot account for all effects of recombination.
We would like to introduce a fast algorithm that reconstructs ultrametric-trees while explicitly accounting for recombination. Instead of considering individual positions of genome sequences, we use a coarse-graining approach, which divides a genome sequence into short segments to account for local density of nucleotide-substitution. For each genome pair considered, our algorithm enumerates the pairwise SNPs on each segment to obtain the pairwise SNP distribution; we fit each empirical SNP distribution to a theoretical SNP-distribution. We test the accuracy of our algorithm against other state-of-the-art algorithms on simulated and real genomes. For genomes with a substantial level of recombination, such as *E. coli*, we show that the age prediction of internal nodes by our algorithm is more accurate than the others, while the tree topology is at least as accurate. Thus, our algorithm is more accurate and faster than alternative recombination-aware methods for ultrametric phylogenetic reconstructions.

Expanded summary*: Horizontal gene transfer (HGT) is a mechanism that allows prokaryotes to exchange DNA segments. After a foreign DNA segment enters a prokaryotic cell, a possible consequence is homologous recombination, where this incoming segment finds a segment on the host genome with high sequence similarity and overwrites it. Thus, homologous recombination erases phylogenetic history, and phylogenetic reconstruction without taking recombination into account can lead to underestimation of the age of ancestral nodes.

The effects of prokaryotic recombination on phylogenetic reconstruction are complicated, and there are multiple factors to consider. Recombination can disturb phylogenetic reconstruction: on the one hand, (i) recombination between the considered genomes can erase nucleotide polymorphisms, making the divergence time of genome pairs appear shorter than they actually are; on the other hand, (ii) recombination of one of the considered genomes with a DNA segment from an outgroup genome—a genome whose lineage has diverged before the most recent common ancestor of the considered genomes has split—will introduce new polymorphisms. The best ultrametric phylogenetic reconstruction algorithms currently available that takes recombination into account, such as ClonalFrame or Bacter package of BEAST2, consider genomic segments with high numbers of nucleotide substitutions to be the results of recombination, and thus accounts for type (ii) recombination; however, it cannot account for type (i) recombination.

In this work, I developed a coarse-graining approach to phylogenetic reconstruction, which is fast and can directly account for homologous recombination. Unlike conventional algorithms that consider every single nucleotide (or amino acid) site in a sequence, this algorithm divides a genome sequence into equal-sized segments. For every genome pair, it considers

the number distribution of sites with non-identical nucleotide on the corresponding segments. The algorithm fits the empirical pairwise distributions to theoretical distributions that combine vertical inheritance with recombination, and thereby infers the ultrametric phylogenetic tree of the genomes considered.

I performed extensive analyses to test this coarse-graining algorithm, using simulated genomes as well as real *E. coli* genomes. For simulated genomes, I performed phylogenetic reconstructions with this coarse-graining algorithm, along with other state-of-the-art algorithms, including RAxML, BEAST and ClonalFrame. For genome populations with a substantial level of recombination, I demonstrated that (i) the node ages predicted by the new algorithm are significantly more accurate than those predicted by alternative methods, and (ii) the topologies of phylogenetic trees predicted by this algorithm are at least as accurate as another tested algorithm.

I performed phylogenetic reconstruction on real *E. coli* genomes using this algorithm, and also using BEAST and ClonalFrame. I evaluated the accuracy of a phylogenetic tree by comparing the tree with the phylogenetic signal inferred from gains and losses of genes in the genomes, assuming that a “species tree” topology resulting in a lower number of inferred gene gains and losses indicated a more accurate phylogenetic tree. I demonstrated that the trees reconstructed by the proposed algorithm are significantly more consistent with the phylogenetic signal inferred from gains and losses of genes than trees reconstructed by other algorithms.

In sum, I believe that this coarse-graining algorithm can contribute to resolving the divergence time of prokaryotic strains by explicitly accounting for the effects of recombination, thereby advancing our understanding of prokaryotic evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

POB-67

Hierarchical Partition Analysis: a fast pattern recognition approach to phylogenomics

Guy Hoelzer [1,*], Rich Drewes [2]

[1]Biology, [2]Biomedical Engineering Graduate Program, University of Nevada Reno, Reno, United States

Abstract: HPA is a novel framework for estimating phylogenetic trees, which is efficient enough to handle genomic-scale datasets. As a parameter-free method, HPA does not rely on estimated parameter values or the use of max/min criteria. It also does not assume a particular model of the evolutionary process, and it yields a rooted tree directly without the need to apply an error-prone rooting procedure.

HPA first conducts a count of taxonomic partitions observed in the data. Variable sites divide the taxa into two or more subsets, which we call 'partitions'. While many of these partitions ultimately correspond to clades, others do not. HPA next attempts to identify the true clades among these partitions by analyzing frequencies of taxon co-occurrences within the partitions, as follows.

1) For each taxon, HPA constructs a ranked list of co-occurrence counts with each other taxon among all observed partitions as a first step.
   1. a. 'Long branch' taxa in the True tree, which introduce excessive homoplasy into the analysis, exhibit similar levels of co-occurrence with each other taxon. These taxa may be removed from the analysis when they occur to maximize accuracy of the inferred HPA tree.
2) In each focal taxon's ranked list of co-occurrences, the drop in number is then calculated from the first taxon to the second, from the second to the third, and so forth, down to the bottom of the list. This creates a list of 'drops'.
3) Co-occurrence count drops are then transformed into percentage drops. At this point each taxon is associated with an ordered list of percentage co-occurrence drops matching the order of the co-occurrence counts. Each percentage drop is associated with a set of taxa above that drop in the ranked list, which corresponds to a particular partition of taxa. When such a partition matches a True clade it is reflected in a relatively large percentage drop, because synapomorphies subtending that clade increase the counts of that partition in the data. Each partition marked by a drop in these lists usually has multiple independently computed percentage drops, one from the perspective of each source taxon, and these are averaged to give a single number for each putative clade.
4) The averaged percentage drops for all partitions indicated in these lists are then combined and ranked in a master list. HPA accepts the partition on the top of the master list as a clade in the HPA tree, and works its way down the list. Any partition that is hierarchically incompatible with those already accepted in the HPA tree is rejected, and the algorithm moves down the master list until the tree is complete or the list is exhausted.
5) HPA is amenable to an efficient bootstrapping algorithm.

HPA also lends itself to inferences about reticulations in the primary HPA tree. When the partitions accepted into the primary HPA tree are removed from the master list, alternative clades generated by reticulate gene flow rise toward the top of the list for the same reason that the primary tree clades rise to the top in the first pass. The potential to explore the evidence of reticulation is an important and unusual feature of HPA. We will show tests and applications of HPA in this presentation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

POB-423

TRAVERSING PHYLOGENETIC HYPOTHESIS SPACE FOR FINDING A BACKBONE OF AN INSECT RELICT GROUP (INSECTA: ODONATA)

Anton Suvorov [1,*], Stanley Fujimoto [1], Paul Bodily [1], Mark Clement [1], Seth Bybee [1]

[1]Brigham Young University, Provo, United States

Poster: Odonata (dragonflies and damselflies) together with Ephemeroptera (mayflies) represent the most ancient insect lineages that developed wings and were capable of active flight. Family-level phylogenetic relationships of odonates remain poorly understood and require extensive revision in light of large “omic” data. Here we analyzed RNA-seq datasets derived from 83 species in order to reconstruct a robust phylogenetic backbone for the order. The data were explored from different angles within each phase of phylogenetic inference. Each of these phases have been shown in multiple previous studies to affect final evolutionary hypotheses. First, three different homology assessment approaches were evaluated to measure direct influence of identified orthologous cluster types (DNA vs. Protein), amount of data and possibility of inclusion of paralogous genes on phylogenetic reconstruction. Second, various combinations of alignment trimming procedures, alignment types, partitioning schemes and tree building approaches (superalignment vs. coalescence based supertree) were compared to identify possible inconsistencies. Using a comprehensive set of fossil calibration points, we dated the final phylogenetic tree and estimated substitution rates for different odonate clades. Together with the calculated missing data and taxon “rougeness”, the substitution rates were investigated to find impact on nodal support inflation. To date , this is the largest odonate dataset analyzed and provided stable recovery of the phylogenetic backbone, although it failed to robustly resolve certain fast-evolving lineages.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

POB-59

A two-state model of tree evolution and its applications to Alu retrotransposition

Niema Moshiri*, Siavash Mirarab

Abstract: Models of tree evolution are essential in reconstructing and interpreting phylogenetic trees. While many such models have been developed, most have been designed to study the evolution of species. However, phylogenies can also be used to study the evolution of genomic elements, such as repeats and duplicated genes. Roughly 10% of the human genome consists of Alu repeats. We introduce a new model of tree evolution called the dual-birth model that captures evolutionary processes specific to these short repeats. Our model extends the traditional birth-only model and allows each evolving entity to be either active, and thus able to propagate, or inactive. The model is characterized by two rates, and adjusting the ratio of the rates controls the tree balance. We present several theoretical results under this dual-birth model, introduce sampling and parameter estimation algorithms, and study the properties of the model in simulations. We then use the dual-birth model to estimate the number of active Alu elements and their rates of propagation and activation in the human genome based on a large phylogenetic tree that we build from close to one million Alu sequences.

Expanded summary*:

Model: We describe a novel model of tree evolution that extends the widely used Yule model to operate with two rates. Evolving elements, for example, Alu repeats, can be either active or inactive. Active elements propagate and create offsprings at a high rate (λa) while inactive elements remain inactive until they active, with a rate much lower than the activation rate ($\lambda b \ll \lambda a$). Once activated, these elements continue propagating with the λa rate. The ratio $r = \lambda a$ controls the balance of the tree and can generate trees as balances λb as Yule with $r = 1$, or completely unbalanced trees with $r = 0$. The model is appropriate for evolving systems such as Alu repeats where some elements propagate while most don't. Such systems would be expected to result in very unbalanced trees, as our model allows.

Probability distributions: We describe algorithms to sample the distribution defined by the model conditioning on a desired number of leaves, n. We then mathematically derive the following probability distributions and expected tree shape properties conditioning on a fixed n: (i) probability distribution over ordered ranked, unordered ranked, and unordered unranked tree topologies, (ii) probability distribution over ordered and unordered trees with branch lengths, (iii) expected number of cherries as $n \to \infty$, (iv) expected fraction of active and inactive elements as $n \to \infty$, (v) expected branch length as $n \to \infty$. We show how these properties, especially the expected branch length and cherry fraction, can be used to estimate parameters of the model given an estimated tree. Importantly, we describe ways to account for estimation error in the phylogenetic analysis.

Simulations: We perform a series of simulations based on the dual-birth model to address two questions: (i) Is the standard maximum likelihood (ML) phylogenetic inference robust to unbalanced trees that can be generated under the dual-birth model? (ii) Can parameters of the dual-birth model be accurately estimated from sequence data using phylogenetic methods?
Our simulation results show that when trees are generated using $r \ll 1$ (i.e., extremely unbalanced trees), using ML to reconstruct the tree can result in a remarkably high topological error. Increasing r (and thus, tree balance) dramatically improves tree accuracy, implying that ML methods perform poorly only when trees deviate from Yule. Only if we increase sequence length dramatically (e.g., to 4, 800bp), topological error decreases for unbalanced trees. Unlike topology, average branch lengths estimated by ML remains relatively accurate across different levels of tree imbalance as well as across different levels of alignment sequence length.
Since ML methods infer grossly overly-balanced trees as true trees deviate from Yule, naive methods of parameter estimation that depend on measures of tree balance may fail. However, we devise a novel technique based on statistical measures of support to improve estimates of model parameters (r, λa, and λb).

Alu analyses: We then use the dual-birth model to study human Alu sequences, a class of Short Interspersed Nuclear Elements (SINEs), each approximately 300 bp long, that exist in the genomes of supraprimates and are often the subjects of phylogenetic analyses in hopes of gaining insight into the history of supraprimate evolution. There are approximately one million Alu elements in the human genome, comprising roughly 10% of the human genome.
Unlike these past studies, because of recent advances in ultra-large MSA, we perform a full phylogenetic analysis of a complete set of 885,011 full-length Alu sequences. We study two questions: 1) How many Alu elements are active? At what rates do inactive Alu elements become active and active elements propagate? To answer these questions, we build an MSA of our set of Alu sequences and infer an ML tree from the MSA. Then, using the ML tree with our aforementioned methods of dual-birth parameter estimation, we estimate multiple biologically relevant parameters of Alu evolution.
We estimate the percentage of Alu elements that have been historically active to be approximately 2%>6.8%, depending on the exact settings used. Our results photogenically confirm a long line of evidence that most Alu elements are not capable of retrotransposition. We further estimate that there are $\lambda a = 1.331 \times 10{-}8$ activation events per year and $\lambda b = 2.475 \times 10{-}6$ propagation events per year, meaning Alu elements become active with a rate of roughly once every 75 million years, and once active, they propagate with a rate of roughly once every 400 thousand years.

Significance: Although the motivation behind the creation of the Dual-Birth model was to improve our studies of Alu transposable elements, the model can be used to study any nucleic element that replicates in a similar fashion. For example, it is known that the evolution of retrotransposons like Alu elements is similar to the evolution of many retroviruses (e.g. Lerat et al. 1999). With the formal definition of the Dual-Birth model as well as the various probability distributions and expectations of tree properties we have defined, the model could be applied to the study of retroviral evolution, and insights gained by constructing an accurate tree using Dual-Birth as a prior on tree distributions or by estimating Dual-Birth parameters (and thus propagation and activation rates) could prove useful in the development of targeted therapeutics. Further, even though the current model assumes constant activation and birth rates across the entire tree, which is an unrealistic assumption for the evolution of certain retroviruses (e.g. HIV, Lemey et al. 2006), our work on the Dual-Birth model could serve as the base for future extensions that allow for variable rates through time, which would be useful the study of HIV evolution. In short, there exist many biological entities that evolve in a two-state manner similar to Alu elements, and the Dual-Birth model can be broadly applied to aid in the study of those entities as well.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

POB-63

Phylogeny-aware, genome-scale reconstruction of the last bacterial common ancestor

Gareth Coleman [1,*], Gergely Szöllősi [2], Tom Williams [1]

[1]School of Earth Sciences, University of Bristol, Bristol, United Kingdom, [2]MTA-ELTE Lendület (Momentum) Evolutionary Genomics Research Group, Hungarian Academy of Sciences, Budapest, Hungary

Abstract: A rooted tree of Bacteria is essential to infer the gene content and reconstruct the metabolism of the last bacterial common ancestor (LBCA). Knowledge of the LBCA's metabolism has implications for reconstructing the metabolism of the last universal common ancestor (LUCA) and testing hypotheses about the early evolution of life on Earth. Many current ideas pertaining to the nature of the earliest life are informed by hypotheses of prokaryotic phylogeny. However, rooting the tree of bacteria has proven difficult, with conventional rooting using outgroups leading to many differing root positions. Recent discoveries of a huge diversity of new uncultured phyla, in particular the Candidate Phyla Radiation (CPR), have further complicated matters, and the relationships between the major bacterial phyla still have little resolution. Here, we attempt to construct a rooted tree of bacteria using probabilistic gene tree-species tree reconciliation methods. These are hierarchical models in which HGTs, gene duplications and losses are integrated into an overall model of genome evolution using amalgamated likelihood estimation (ALE), and in which patterns of gene family evolution contain information about the root of the tree. We use this rooted tree in order to reconstruct the metabolism of the ancestral bacterium, and compare our results to those obtained from the use of these methods on Archaea, which allows us to infer information on the nature of LUCA. We also use ALE to investigate the rate of bacterial HGT over time, to evaluate how HGT has affected vertical phylogenetic signal and the evolution of early cells.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

POB-66

Species tree, frequent HGT and gene conversions in Bdelloid Rotifers

Marie Cariou [1,*], Hélène Henri [2], Nicolas Debortoli [1], Matthieu Terwagne [1], Boris Hespeels [1], Julie Virgo [1], Damien De Vienne [2], Karine Van Doninck [1]

[1]Unité de Recherche en Biologie Environnementale et Evolutive, Unamur, Namur, Belgium, [2]Laboratoire de Biométrie et Biologie Evolutive, UMR 5558, Université Lyon 1, Villeurbanne, France

Abstract: The vast majority of animals reproduce sexually, i.e. with recombination of genetic material between generations. This genetic mixing is thought to favor the persistence of sexual lineages despite a theoretical advantage of asexuals regarding colonization capacity and population growth. In this context Bdelloid Rotifers, a highly diversified group of animals evolving asexually for millions of years, appears as a puzzle but also as an outstanding model to better understand long-term evolution in the absence of sexual reproduction.

The sequencing of the first Bdelloid Rotifer genome (Flot et al. 2013) revealed a peculiar organization, characterized by a degenerate tetraploidy, numerous rearrangements and the absence of homologous chromosomes. This structure, incompatible with meiosis, likely testifies the long asexual evolution of this genome. Besides, population genetic studies suggest that Bdelloids might exchange DNA within and between species, which likely plays a major role in their evolutionary history (Debortoli et al. 2016).

To better understand the dynamics of gene conversion and horizontal genetic transfers on these asexually evolving genomes and their importance in the success of Bdelloids, we are generating genomic data (RAD sequencing) from a large number of species distributed across the four existing families. We will show how methods involving the reconciliation of gene and species trees can help resolving species phylogeny of Bdelloids despite asexuality, polyploidy, gene conversions, transfers and losses. The resolution of the species tree is indeed a requirement for further study of the dynamics of horizontal transfers, which might play a key role in their evolution.

Expanded summary*: The vast majority of animals reproduce sexually, i.e. with recombination of genetic material between generations. This is despite a theoretical advantage of asexuals regarding colonization capacity and population growth and thus, it prompted the hypothesis that recombination of polymorphic loci allowed by sexual reproduction might favor the persistence of sexual lineages. In this context Bdelloid Rotifers, a highly diversified group of animals evolving asexually for millions of years, appears as a puzzle but also as an outstanding model to better understand long-term evolution in the absence of sexual reproduction.

The sequencing of the first Bdelloid Rotifer genome (Flot *et al.* 2013) revealed a peculiar organization, characterized by a degenerate tetraploidy, numerous rearrangements and the absence of homologous chromosomes. This structure, incompatible with meiosis, likely testifies the long asexual evolution of this genome. Besides, population genetic studies suggest that Bdelloids might exchange DNA within and between species, which likely plays a major role in their evolutionary history (Debortoli *et al.* 2016).

Our project aims at a better understanding of the importance of these genetic transfers in the success of Bdelloids. We are generating genomic data (RAD sequencing) from a large number of species sampled in the wild and distributed across the four existing families. This data will be useful to investigate the genomic history and dynamic of Bdelloids genomes. However, clonal lineages and partial polyploidy, associated with a significant rate of gene conversions and potential horizontal transfers might hinder the resolution of species phylogeny. We will show how methods involving the reconciliation of gene and species trees can help resolving the species phylogeny of Bdelloids despite asexuality, polyploidy, gene conversions, transfers and losses.

The resolution of the species tree is indeed a requirement for further study of the dynamics of horizontal transfers, which might play a key role in Bdelloid evolution. It will allow quantifying recombination between lineages, at different time scale, and thus provide insights on the genomic and ecological determinant of these transfers. Do gene transfers mainly occurs within closely related genetic clusters or also on a larger scale as is suggested by the high percentage of horizontally acquired genes non-metazoan origin in *Adineta vaga*'s genomes? Another striking characteristic of Bdelloids is their ability to withstand desiccation at any life stages. Such desiccation events have been shown to result in numerous double strand breaks in genomes, which are subsequently repaired (Hespeels *et al.* 2014). This observation led to the hypothesis that these events might favor the integration of foreign DNA in Bdelloid genomes. Thus, a better understanding of the dynamic and evolutionary importance of horizontal transfers between Bdelloid's

genomes will be complementary to studies allowing a direct measure of the impact of desiccations on horizontal transfers between rotifers.

Keywords:
Bdelloid rotifers, sex evolution, phylogenetics

References:

Debortoli N, Li X, Eyres I *et al.* (2015) The end of an ancient asexual scandal : sapphomixis in the bdelloid rotifer Adineta vaga. *Current Biology*, **27**, 1334–1345.
Flot J, Hespeels B, Li X *et al.* (2013) Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga. *Nature*, **500**, 453–457.
Hespeels B, Knapen M, Heuskin A, Pineux F (2014) Gateway to genetic exchange ? DNA double-strand breaks in the bdelloid rotifer Adineta vaga submitted to desiccation. *Journal of Evolutionary Biology*, 1–12.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

POB-65

Exploring the evolution of visual opsins within the Ecdysozoa using wider genomic sampling and new lineage-specific replacement-rate models.

James Fleming [1,*], Davide Pisani [1], Nicholas Roberts [2]

[1]School of Earth Sciences, [2]School of Biological Sciences, University of Bristol, Bristol, United Kingdom

Abstract: The Ecdysozoa - the Scalidophora (Priapulida, Kinorhyncha, Loricifera), Nematoida (Nematoda, Nematomorpha) and Panarthropoda (Tardigrada, Onychophora and Arthropoda) –is a useful model system of evolution, as it includes both highly diverse clades such as the Arthropoda and depleted phyla like the Priapulida - for which only eight genera are known. Currently available genomic information for the Ecdysozoa, however, almost exclusively focuses on the Arthropoda and Nematoda.
The Arthropoda and Vertebrata are the only known phyla to have developed colour vision, which in the arthropods is achieved through the use of multiple different rhabdomeric opsin proteins tuned to different wavelengths of light. The number of visual opsins present in any given arthropod species can vary greatly, from none to more than thirty. Colour vision is thought to be the basal state for the Arthropoda, but almost nothing is known of the visual history or opsin distribution of the other ecdysozoan clades.
We present the first evidence for the existence of visual opsins across all major ecdysozoan lineages – the Scalidophora, Nematoida and Panarthropoda. Using cutting-edge phylogenetic approaches (including lineage specific replacement-rate heterogeneous models) we show that multiple independent visual opsin duplications have occurred across the Ecdysozoa, not only in the Arthropoda. With the use of molecular clock approaches we have shown that the emergence of Arthropod colour vision can be dated to between 634 and 595 MA.

Expanded summary*: The Ecdysozoa are an important group of animals, containing the majority of named species and both hyper-diverse clades such as the Arthropoda, and heavily depleted clades like the Priapulida. The Ecdysozoa comprises the Priapulida, Loricifera, Kinorhyncha, Nematoda, Nematomorpha, Tardigrada, Onychophora, Chelicerata, Myriapoda, Crustacea and Insecta; outside of the arthropods, however, the vision of these creatures has been relatively unexplored, and no studies have assessed the visual capabilities of the Priapulida, Loricifera, Kinorhyncha and Nematomorpha.
Vision is a photoreceptive technique that allows organisms to process light information beyond a simple photic response. From a molecular perspective, vision in the Ecdysozoa is regulated by rhabdomeric (R) opsins, a form of light sensitive protein that is expressed within the eyes of the organism. Expression of multiple opsins at the same time, tuned to different wavelengths, and the capability to differentiate between different responses from the different opsins is what enables colour vision, a state that is thought to have arisen only twice: once in the Arthropoda, and once in the Vertebrata – though the Vertebrata utilise a system based on ciliary opsins.
Ecdysozoan colour vision is incredibly important when considering evolutionary history and ecology. The ability to perceive one's environment is closely linked to the needs and requirements of that environment, and so understanding when and under what ecological and historical pressures opsin gene duplications occurred – enabling organisms to respond to an additional, different set of wavelengths – is important to understanding the development of life on earth.
Before this study began, due to the poor sampling of the Tardigrada (1 species, *Hypsibius dujardini*), and Myriapoda, the history of colour vision in the Arthropoda was still unclear. A complete lack of sampling in any clades of the Ecdysozoa more basal than the Tardigrada also needed to be remedied to understand the history of vision.
In addition, solving a current phylogenetic controversy helped motivate this research. The invertebrate non-visual r-opsins, or Arthropsins, are a group of r-opsins that are often recovered basal to the vertebrate non-visual r-opsins (melanopsins). This position requires a considerable number of independent opsin gene losses in the Lophotrochozoa, Arthropoda and Vertebrata. To help resolve this problem, we have created independent rate matrices for the opsins that outperform the GTR and WAG models in both likelihood and Bayesian scenarios, and explored lineage-specific heterogeneous modelling (a relatively new phylogenetic method whereby substitution matrices are assigned to particular branches of a scaffold tree topology) as a way to resolve contentious phylogenetic problems.

This study presents new transcriptomic and genomic information from the Priapulida, Kinorhyncha, Nematomorpha and Tardigrada. The results have shown that genera-specific duplications in the Tardigrada appear to be relatively common, and confirm that colour vision is specific to the Arthropoda. It also shows that the loss of opsins is specific to the Nematoda, not the Nematoida, as previously thought. In timing the origin of colour vision, we can suggest a number of potentially realistic ecological causes for such an event, including an exploration and colonisation of new environments, that gives useful context to existing fossil evidence. Finally, new ecdysozoan r-opsin sequences have proposed a new, more parsimonious position for the arthropsin clades.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

OW-CH6

Reconstructing species substitution rates and divergence times from nuclear DNA using StarBEAST2

Huw Ogilvie [1 2,*], Craig Moritz [1], Alexei Drummond [2 3]

[1]Research School of Biology, Australian National University, Canberra, Australia, [2]Centre for Computational Evolution,

[3]Department of Computer Science, University of Auckland, Auckland, New Zealand

Abstract: Relaxed clocks can be used to estimate variation in substitution rates among extant and ancestral species, but existing computational models do not fit the best sources of data. Mitochondrial genome alignments are often used to estimate substitution rate variation, however mtDNA may be evolving too quickly (in animals) or too slowly (in plants) to contain enough information at the divergence time scales researchers are interested in. Nuclear genomes are a much larger and richer source of information, however unlinked nuclear segments do not share a common gene tree due to recombination, and because relaxed clocks have until now assumed a common gene tree this leads to systematic errors when inferring substitution rates and divergence times. We introduce species tree relaxed clock models, which model per-species substitution rates but also allow for many gene trees embedded within a species tree. This extends the multispecies coalescent model of species and gene evolution, and enables more accurate inferences in studies that incorporate substitution rates, such as divergence dating or studies of tempo and mode. Species tree relaxed clocks are implemented in the StarBEAST2 package for BEAST2. We will demonstrate the importance of species tree relaxed clock models for reliable substitution rate and time estimates by applying StarBEAST2 to exon-capture sequence alignments from Eugongylus group skinks.

Expanded summary*: Substitution rates are of interest to researchers working on molecular drivers of speciation and adaptation. They are also highly correlated with divergence time parameters, so accurate dating depends on accurate estimates of substitution rates. For a given species, the average substitution rate is correlated with a multitude of traits including metabolic rate, body size, and fecundity (see Bromham, 2011 “The genome as a life-history character”). This implies that when studying clades with variation in those traits, substitution rates will vary between species.

Variation in substitution rates can be accommodated in concatenation analyses. However substitutions produced by incomplete lineage sorting (SPILS) causes concatenation to overestimate the lengths of specific branches and underestimate the lengths of others, which produces apparent substitution rate variation where none exists (see Mendes and Hahn, 2016 “Gene tree discordance causes apparent substitution rate variation”). We show that this translates into systematic bias when estimating divergence times and substitution rates.

Unlike concatenation, the multispecies coalescent (MSC) models multiple gene trees evolving within a species tree. Because variation is expected in the relative substitution rate of different genes and different species, MSC models should take both per-gene and per-species rate variation into account. We have developed species tree relaxed clock models that for the first time extend the MSC to apply and a separate substitution rate for each species, in addition to a separate substitution rate for each gene.

We have implemented multiple species tree relaxed clock models in the StarBEAST2 package for MSC inference, in addition to vast improvements in computational performance. These include uncorrelated log-normal clocks, uncorrelated exponential clocks, and random local clocks. StarBEAST2 will thereby enable reliable estimates of divergence times and substitution rates despite gene tree discordance.

Armed with accurate dating of divergence times, we can begin to understand the context in which speciation events occurred. This helps us to answer why we and other organisms are here on Earth today, a question of great general interest based on coverage of evolutionary findings in popular media. It also assists scientists and governments conserve biodiversity, for example by accurately identifying areas of strong or weak phyloendemism across a landscape.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

POB-57

Impact of long-term chromosomal shuffling on multispecies coalescent analysis in two lineages of anthropoid primates

Carlos Schrago [1,*]

[1]Genetics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Abstract: Phylogenetic inference has become theoretically unified with population genetics theory following the introduction of multispecies coalescence (MSC) analysis for estimating species trees. However, despite successful MSC applications, this method has been questioned in favor of the standard concatenation approach to phylogenetics because it assumes that gene trees inferred from individual loci represent independent trials of the MSC process. Since genes might be physically close to one another within syntenic associations spanning along chromosome regions of different size, fundamental MSC theoretical assumptions on the individuality of loci might be flawed when analyzing evolutionary lineages with accelerated rates of chromosome evolution and substantial karyotypic shuffling. In this respect, neotropical primates represent an ideal case for assessing the performance of MSC methods because chromosome diploid number varies significantly in this lineage. Here, we investigated the effect of sequence size on the performance of coalescent-based methods. This analysis was carried out by comparison with the hominid (great apes and human) lineage in which chromosome macrostructure has been stable during 15 million years of evolution. We found that both heuristic MSC methods and concatenation performed well, showing statistical consistency and robustness for inferring the correct species tree. Departure from the MSC model in neotropical primates was reduced with smaller sequence fragments where coincidental phylogenetically informative sites were found more frequently than in longer fragments. This picture may have resulted from extensive karyotypic rearrangement occurring during the radiation of neotropical primates contrary to the comparatively stable chromosome evolution in the hominids.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

OW-CH4

A timescale for life's early evolution: Is it time to leave behind a literal interpretation of the fossil record?

Holly Betts [1,*], Tom Williams [1], Philip C. J. Donoghue [1], Davide Pisani [1]

[1]University of Bristol, Bristol, United Kingdom

Abstract: The timescale of life's early history on Earth remains one of the last areas where literal interpretations of the fossil record guide our understanding. However, the uncertainty surrounding the early fossil record of life suggests that a probabilistic timescale could serve as better guideline. Such a timescale can be generated integrating fossils and genomic information using modern, relaxed, molecular clock methods. Molecular clocks rely on carefully constructed calibrations and, to date, no definitive set of calibrations for dating fundamental divergences within the tree of life has been assembled. Here, we establish a suite of calibrations, employing them with the molecular clock to show that divergence times are sensitive to calibration distribution choice, and the clock model used. Integrating across the uncertainties yields a timescale that defines credibility intervals for key events in the history of life. Our probabilistic timescale, integrating fossil and genomic information, is more accurate though less precise, allowing for predictive power and simple refinements as new fossil and molecular data are revealed. Eukaryotes emerge late in the history of life (~1.54) Ga, in good agreement with the known fossil record. Crucially, the Alphaproteobacteria (the clade to which the mitochondria belong) also emerge in the same age range (0.89-1.57), confirming a fundamental role for symbiosis in the establishment of the eukaryotic lineage.

Expanded summary*: The premise of our work was to try and date the early divergences of life, and some of the key evolutionary splits, such as that of the eukaryotes from the Archaea. The fossil record at the time that these divergences were likely occurring is sparse, owing to a combination of rock loss and that any rock remaining has often been highly metamorphosed. On top of this there is considerable controversy about the biological authenticity of the oldest fossils, as each one has few characteristics to link it to any distinct group. Despite the lack of evidence each new fossil is used to reinterpret the history of life at this time, an approach that is no longer used in more recent timescales due to the application of molecular clocks, a statistical approach to integrate over the uncertainty found in the fossil record. Ironically these clocks have not yet been applied to the deepest nodes of life, where their use may be most relevant.

Initially we collated two datasets, one of genetic material, and one of fossil information to produce the calibrations for the molecular clock process. Fossils are imperative here in that they help to produce dates in absolute time. Our genetic dataset consists of 29 proteins that are common to all 3 domains of life. They are all involved in necessary functions, such as within the ribosome, and are highly conserved and therefore useful for dating a very diverse range of species. The calibrations are mostly within the Eukaryotes, as they have the most complete fossil record, and one that is more easily identified based on external characters. One of the most important calibrations is on the root node, constraining not only LUCA but the whole tree of life. For this we used the moon forming impact, an event so powerful that it would have reformed, and in so doing sterilised the Earth. Hence, the last universal common ancestor (LUCA) must have existed afterwards.

Our results date LUCA to ~4.4 Ga and show the crown prokaryote lineages diverging over one billion years after this, at similar times, in the Archaean ~3 Ga. This date for LUCA is much older than the oldest potential fossil records at Isua, Greenland dated to around 3.8 Ga. The concurrent emergence of the Bacteria and Archaea is interesting as today the diversity of the Bacteria far outweighs what we see in the Archaea and yet they have much the same evolutionary timescale. The crown eukaryotes diverge much later in the mid Proterozoic, around the time we might expect, based on the fossil record, and at a comparable date to other studies that place the eukaryotes emergence at around 1.2-1.8 Ga. Although the eukaryote tree has been dated multiple times before and some work has been carried out on dating bacterial lineages, the tree of life as a whole has not previously been tackled. One of the most important geological events, the great oxidation event, which has been linked to both the cyanobacteria and the eukaryotes, occurs before the evolution of both of these crown groups in our study. This is contrary to most reports in which the cyanobacteria are causally associated with the GOE.

Our investigation confirms that the oldest fossil can never give us the true age of a clade, but when integrated with molecular data can be used to produce a robust timescale for the tree of life. It is a framework that can be built upon and updated as new fossils are discovered and new genetic sequence material becomes available.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

OW-CH3

Phylogenetic rooting using minimal ancestor deviation

Fernando Tria [1,*], Giddy Landan [1], Tal Dagan [1]

[1]Institute of General Microbiology, University of Kiel, Kiel, Germany

Abstract: Ancestor-descendent relations play a cardinal role in evolutionary theory. Those relations are determined by rooting phylogenetic trees. However, existing rooting methods are hampered by evolutionary rate heterogeneity or the unavailability of auxiliary phylogenetic information. We present a novel rooting approach, the minimal ancestor deviation (MAD) method, which embraces heterotachy by utilizing all topological and metric information in unrooted trees. We demonstrate the method in comparison to existing rooting methods by the analysis of phylogenies from eukaryotes and prokaryotes. MAD correctly recovers the known root of eukaryotes and uncovers evidence for cyanobacteria origins in the ocean. MAD is more robust and consistent than existing methods, provides measures of the root inference quality, and is applicable to any tree with branch lengths.

Expanded summary*: Phylogenetic tree reconstruction methods produce unrooted trees, which require an additional analysis – rooting – in order to determine the ancestor-descendant relations of the studied entities. Given its pivotal role for evolutionary studies, it is notable that no major advance in the methodology of rooting has been presented since the publication of the leading rooting approaches in the 1970's (i.e., outgroup and midpoint). This situation is in stark contrast to the wide range of methods available for the reconstruction of phylogenetic tree topologies. We introduce the Minimal Ancestor Deviation (MAD) rooting method, which operates on any phylogenetic tree with branch lengths. Our method explicitly quantifies the major confounding factor - rate variation among lineages - to achieve unprecedented accuracy and consistency. We anticipate that many long-standing evolutionary controversies will be settled (or at least sharpened) with high-quality rooting. An example is our rooting of the cyanobacteria, showing that the basic photosynthetic machinery originated in a marine environment - at a consistency level of 69%(!) of the gene families. Such a consensus is rarely encountered in analyses of deep prokaryotic phylogenies, which notoriously suffer from reticulated evolution and reconstruction errors. Rooting, and the resolution of ancestor-descendant relations, is by no means solely in the domain of evolutionary biology. Epidemiological reconstruction of infection trajectories, ancient DNA studies and linguistic cognate reconstruction are among the fields that utilize trees and will benefit from a versatile and reliable rooting method.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

POB-64

Fast and Accurate Estimates of Divergence Times from Big Data

Beatriz Mello [1,*], Qiqing Tao [2], Koichiro Tamura [3], Sudhir Kumar [2]

[1]Department of Genetics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, [2]Department of Biology, Temple University, Philadelphia, United States, [3]Department of Biological Sciences, Tokyo Metropolitan University, Tokyo, Japan

Abstract: Ongoing advances in sequencing technology have led to an explosive expansion in the molecular data available for building increasingly larger and more comprehensive timetrees. However, Bayesian relaxed-clock approaches frequently used to infer these timetrees impose a large computational burden and discourage critical assessment of the robustness of inferred times to model assumptions, influence of calibrations, and selection of optimal data subsets. We analyzed eight large, recently published, empirical datasets to compare time estimates produced by RelTime (a non-Bayesian method) with those reported by using Bayesian approaches. We find that RelTime estimates are very similar to Bayesian approaches, yet RelTime requires orders of magnitude less computational time. This means that the use of RelTime will enable greater rigor in molecular dating, because faster computational speeds encourage more extensive testing of the robustness of inferred timetrees to prior assumptions (models and calibrations) and data subsets. Thus, RelTime provides a reliable and computationally thrifty approach for dating the tree of life using large-scale molecular datasets.

Expanded summary*: Progress in sequencing technology has led to a two-dimensional expansion of datasets being used for dating evolutionary divergences, because both the number of sites in the sequence alignment and the number of included taxa are increasing quickly. Large time-calibrated phylogenies are being generated using these data, helping to elucidate the evolutionary patterns and the underlying processes responsible for the extant biological diversity. Recently, Bayesian approaches have been the most commonly used to estimate biological timescales. However, with the rise of big data, the application of Bayesian methods for dating cladogenetic events is becoming computationally demanding.

Their time requirements increase exponentially with increases in the number of species and the sequence length. For example, it takes almost half a day to compute divergence times in a dataset with 43 species (~55k sites); and multiple days to estimate a timetree for 274 mitochondrial sequences (first and second codon positions) on a personal computer (Intel® Core i7® CPU @ 4.0GHz). Such slow speeds slow down the pace of discovery and even lead to suboptimal scientific practices, because they discourage tests of the robustness of inferred timetrees to the model and calibration assumptions.

Recently, ultra-fast non-Bayesian dating methods have been developed, which allow rate variation from branch to branch and incorporate multiple calibration points. These methods have already been shown to produce excellent estimates for simulated data. For example, the performance of RelTime method was comparable to Bayesian approaches in computer simulations where large sequence datasets were generated under conditions with autocorrelation and independent rates among lineages. Also, RelTime method produced estimates that were frequently better than Bayesian and other approaches, especially when there was a 50% rate speedup in a specific clade. Importantly, non-Bayesian approaches complete calculations thousands of times faster than the fastest Bayesian method, with even greater speed differences for larger numbers of sequences.

Therefore, we were prompt to investigate if RelTime produces divergence time estimates that are comparable to those obtained using Bayesian methods on empirical datasets, especially when they are very large. Because if this is true, then RelTime would provide a computationally tractable alternative to Bayesian methods. Thus, we directly compared Bayesian and RelTime methods by reanalyzing eight large-scale empirical datasets obtained from recently published studies. In these sequence alignments, the number of taxa ranged from 36 to 274 and the number of sites ranged from 7,370 to 20,593,949 (nucleotides or amino acids). These datasets represent some of the biggest timetree analyses performed to date.

We found a huge concordance of time estimates between Bayesian and RelTime approaches, which was strong across datasets that vary extensively in numbers of taxa and length of the sequence alignment. Therefore, RelTime provides an accurate and computationally-efficient approach to estimate times when Bayesian methods are infeasible. Furthermore, achieving similar results from two distinct approaches increases our confidence in biological conclusions. As Bayesian methods require many more priors than RelTime, we recommend that RelTime should be applied along with Bayesian and other approaches.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Calibrating the history of life

OW-CH5

Reconstructing a dated tree of life using phylogenetic incongruence

Adrian D Davin [1], Eric Tannier [1], Bastien Bossau [1], Vincent Daubin [1], Gergely Szollosi [2,*]

[1]LBBE, UMR CNRS 5558, Lyon, France, [2]MTA-ELTE "Lendulet" Research Group, Eötvös University, Budapest, Hungary

Abstract: Past evolutionary events recorded in the DNA of living organism might be the key to study the Early History of Life, a long period of time for which the fossil record is scarce and unreliable. The exchange of genes among different species have left tell-tale footprints of phylogenetic incongruence in extant genomes that can be detected with modern phylogenetic techniques. These genomic fossils can tell us which ancient species lived at the same time and thus can be used for dating studies. I describe our recent results that show that transfers detected using gene tree-species tree reconciliations carry a strong time signal resembling those of paleontological fossils across the three domains of life. I also show that different methods for relaxing the molecular clock produce varying fit with the dating information conveyed by transfers, hence transfer events can potentially be used to choose among competing alternatives. Finally, I discuss our plans for developing genome-scale dating methods that exploit the genomic fossils recorded by phylogenetic incongruence in the context of the recently funded ERC project "GENECLOCKS".

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-299

Ancestral sequence reconstruction while accounting for protein structural information

Asher Moshe*, Tal Pupko and Pupko lab

Abstract: Ancestral sequence reconstruction (ASR) was shown be an important tool that helps understand the evolutionary origin of modern proteins. In addition, ancestral proteins often contain desired properties that modern proteins lack, e.g. broader substrate range, therefore they can be used as a good starting point for protein engineering.
FastML is a web tool built to infer ancestral sequences, based on the maximum likelihood paradigm. Although a number of tools for ASR exist, FastML contains several features that differentiate it from the rest e.g. the way it treats gaps and its flexibility regarding the allowed evolutionary models. FastML was shown to be one of the most accurate tools available for ASR in a recent benchmark test.
Currently, most models use a single substitution matrix to generate the ancestral sequences. This approach, while giving relatively satisfying results with high likelihood score, is far from perfect. When working with proteins, there is a large variability between different parts of the protein, especially in terms of solvent accessibility. Therefore, using a single average substitution matrix, e.g. LG, will give less than optimal results. In this work, we hypothesized that using a mixture of amino acid replacement matrices to infer ancestral sequences, would lead to a significant increase in ASR accuracy, and that accuracy would substantially increase when amino-acid replacement matrices would be developed while accounting for structural information when data are available.
We aim to incorporate all these models into the FastML algorithm and test their impact in simulations and on the benchmark dataset.

Expanded summary*: Ancestral sequence reconstruction while accounting for protein structural information

Abstract

Ancestral sequence reconstruction (ASR) was repeatedly shown be an important tool that helps understand the evolutionary origin of modern proteins. In addition, ancestral proteins often contain desired properties that modern proteins lack, such as broader substrate range and higher thermostability, therefore they can be used as a good starting point for protein engineering.[1]
FastML is a web tool built to infer ancestral sequences. It is based on the maximum likelihood paradigm. Although a number of tools for ASR exist, FastML contains several features that differentiate it from the rest e.g. the way it treats gaps and its flexibility regarding the allowed evolutionary models.[2] FastML was shown to be one of the most accurate tools available for ASR in a recent benchmark test.[3]
Currently, most substitution models use a single substitution matrix to generate the ancestral sequences. This approach, while giving relatively satisfying results with high likelihood score, is far from perfect. When working with proteins, there is a large variability between different parts of the protein, especially in terms of solvent accessibility. Therefore, using a single average substitution matrix, such as WAG or LG, will give less than optimal results. In this work, we hypothesized that using a mixture of amino acid replacement matrices (instead of a single matrix) to infer ancestral sequences, would lead to a significant increase in ASR accuracy. We further hypothesized that accuracy would substantially increase when amino-acid replacement matrices would be developed while accounting for structural information assuming such data are available.[4]
Finally, when 3D structure is unavailable, we aim to test if the prediction of structural features (such as buried and exposed positions) and use of structural aware models would increase accuracy, compared to using non-structural aware models. We aim to incorporate all these models into the FastML algorithm and test their impact in simulations and on the recently developed empirical benchmark dataset.[3]

References

1. 1. Gumulya Y. et al. Biochemical Journal (2017) 474 1–19 DOI: 10.1042/BCJ20160507
2. 2. Ashkenazy H. et al. W580–W584 Nucleic Acids Research, 2012, Vol. 40, Web Server issue. doi:10.1093/nar/gks498

3. 3. Randall RN et al. Nature communications 2016 7:12847 | DOI: 10.1038/ncomms12847
4. 4. Le SQ et al. Syst. Biol. 59(3):277–287, 2010. DOI:10.1093/sysbio/syq002

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-300

Development of a RAD sequencing approach to discover and genotype genome-wide genetic markers for New World Primates

Lina Valencia [1], Amely Martins [2 3], Edgardo M Ortiz [4,*], Anthony Di Fiore [1]

[1]Primate Molecular Ecology and Evolution Laboratory, Department of Anthropology, University of Texas at Austin, [2]Primate Molecular Ecology and Evolution Laboratory, Department of Anthropology, University of Texas, Austin, Austin, United States, [3]Centro Nacional de Pesquisa de Conservação de Primatas Brasileiros, ICMBio/MMAm, Brazil, Brazil, [4]Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: Most genetic studies of nonhuman primates have utilized mitochondrial DNA and/or a small number of nuclear DNA markers, which can limit our understanding of primate population genetics and phylogenetic relationships. However, the increasing availability and decreasing cost of "next generation" sequencing technologies now allows the discovery, sequencing, and genotyping of thousands of genetic markers distributed throughout the genome rapidly, at relatively low cost, and without relying on the availability of a reference genome. We employed a "double digest Restriction-Site Associated DNA" (ddRAD-seq) next-generation sequencing method to identify single nucleotide polymorphisms (SNPs) in 12 genera of neotropical primates belonging to the three platyrrhine families, Pitheciidae, Atelidae, and Cebidae. We digested genomic DNA samples with four restriction enzyme pair combinations to find the optimal set to use across a taxonomically diverse sample of platyrrhines. The combination of *Sph*I+*Mlu*CI performed efficiently for all species and generated millions of short paired-end sequence reads on the Illumina HiSeq 4000 platform, from which we identified thousands of SNPs for each of the target taxa. Phylogenetic analyses using high-confidence SNPs resulted in a well-supported phylogeny concordant with other molecular phylogenetic studies of platyrrhines based on targeted sequencing of small numbers of loci. The SNP data also allowed us to make high resolution inferences about the relationships among different populations of select species included in our sample. These results demonstrate the utility and promise of using ddRAD-seq to efficiently and inexpensively discover genome-wide markers in a host of New World primates for which existing genomic resources are lacking.
Supported by the NSF (BSC-1540270), UT Austin, PSGB, IPS, PCI, CI and The Rufford Foundation

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

OTH-CR5

Why versioning is crucial for evolutionary systems biology

Cecilia Moog [1], Jamieson Warner [2], Laurence Loewe [1,*]

[1]Laboratory of Genetics and Wisconsin Institute for Discovery, [2]Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, United States

Abstract: Versioning systems often go unconsidered in biology as the labeling of version variants is typically deemed a trivial task. However, human error analyses tell us to expect an error in every several hundred manual tasks performed, even in the best cases. Biologists handling large amounts of data know how pivotal it can be to distinguish different variants of a dataset as they are being refined over the course of a research project. Storing all data in Versioned Biological Information Repositories (VBIRs) enables precise citations that link to the exact states of respective causal datasets that were used to draw a given consequential conclusion. Such clarity is a necessary condition for reproducibility. For example, the "official" publication of the human genome project presented a mostly complete version that has since been refined and updated numerous times. Such updates may be desirable as (i) technology improves and enables more accurate re-analysis, (ii) errors in a VBIR are discovered and patched, or (iii) other more causal VBIRs are improved. For example, human genome annotation depends on correct gene ontologies. Likewise in evolutionary genetic analyses, changing versions can have a substantial effect on results and biological implications. Versioning problems motivate numerous efforts (e.g. https://galaxyproject.org) to improve efficiency and reproducibility by avoiding the unnecessary complications that arise from miscommunicated version information. A versioning system that is stable across projects could reduce such miscommunication. - Evolutionary Systems Biology (EvoSysBio) requires the integration of large amounts of data for describing the fitness causality networks that underpin fitness landscapes. Such attempts at mapping genotypes to phenotypes in non-trivial systems requires deeply nested causality networks, which are best organized into a network of VBIRs. This high connectivity implies that data errors in one VBIR can have a ripple effect on many others. Correcting such errors is most efficiently done at the source, if an automated mechanism exists for passing on the corrections. Such a mechanism requires a shared way of labeling the reliability and importance of a given version variant to effectively distinguish trivial typo corrections from scientifically relevant changes. It is difficult to achieve such efficient version resolution if every VBIR happens to use a nomenclature for distinguishing version variants that was chosen by historical accident. - Here we present a stabilizing versioned number system designed to facilitate efficient communication of versioning information across systems and research groups. We have been developing a marked positional number system that can automatically generate version variant numbers. To be as stable as required for resolving the versioning problems above, this system needs to address all conceivable use-cases it might encounter in the real world. We have collected use cases through our experience with biological model curation and development of scientific computing software; however, this is by no means an exhaustive list. We invite feedback on use cases unknown to us and welcome thorough reviews of our current design for a stabilizing versioning system (see http://evolvix.org/versioning). More about VBIRs can be found at https://doi.org/10.1101/099192, and reviews of a mechanistic view of EvoSysBio are at http://evolutionarysystemsbiology.org

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

OTH-CR2

The New Advanced Computing Landscape: Opportunities for Molecular Biology

Dan Stanzione [1,*], Matthew Vaughn [1]

[1]Texas Advanced Computing Center, Austin, United States

Abstract: This talk will cover recent developments in the high performance computing ecosystem – in architectures, systems, software techniques, APIs, and tools – and the push within the ecosystem to provide better support for a wide range of computational science activities beyond simulation, including data analysis, reproducibility, collaboration, etc. Examples will draw on a range of new systems at the Texas Advanced Computing Center, such as Stampede-2 (which will enter production just prior to the meeting), and from the Galaxy, iPlant, Cyverse, and DesignSafe software ecosystems.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

OTH-CR6

Infectious Disease Dynamics Inferred from Genetic Data via Sequential Monte Carlo

Alex Smith [1,*], Edward Ionides [2], Aaron King [3]

[1]Bioinformatics, [2]Statistics, [3]Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, United States

Abstract: Genetic sequences from pathogens can provide information about infectious disease dynamics that may supplement or replace information from other epidemiological observations. Currently available methods first estimate phylogenetic trees from sequence data, then estimate a transmission model conditional on these phylogenies. Outside limited classes of models, existing methods are unable to enforce logical consistency between the model of transmission and that underlying the phylogenetic reconstruction. Such conflicts in assumptions can lead to bias in the resulting inferences. We have developed a general, statistically efficient, plug-and-play method to jointly estimate both disease transmission and phylogeny using genetic data and, if desired, other epidemiological observations. This method explicitly connects the model of transmission and the model of phylogeny so as to avoid the aforementioned inconsistency. We demonstrate the feasibility of our approach through simulation and apply it to estimate stage-specific infectiousness in a subepidemic of HIV in Detroit, Michigan. While we focus on how these methods may be applied to population-level models of infectious disease, their scope is more general. These methods may be applied in other biological systems where one seeks to infer population dynamics from genetic sequences, and they may also find application for evolutionary models with phenotypic rather than genotypic data.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-301

EXONtools: A complete pipeline for exon capture sequencing data analysis of non-model organisms

Kirill Vinnikov [1,*], Kathleen Cole [1]

[1]Department of Biology, University of Hawaii at Manoa, Honolulu, United States

Abstract: Exon capture sequencing is one of the most promising approaches in evolutionary and population genomics because it only obtains data from protein coding DNA regions (~2%), avoiding all other unnecessary sites. As a result, hundreds of individuals can be multiplexed and analyzed all together during a single high-throughput sequencing run, yet still providing a substantial amount of data required for deep phylogenomic reconstructions, large scale population studies, identification of candidate genes under selection, etc. Furthermore, in combination with *de novo* transcriptome sequencing (RNA-seq), the difficult task of capturing whole exomes for non-model organisms can also be accomplished. Written in Python programming language, EXONtools software represents a complete pipeline for exon capture sequencing data analysis. It implements several novel bioinformatics algorithms along with the use of existing popular programs into one analytical framework. The pipeline consists of three major steps: 1) transcriptome assembly and annotation; 2) exon bait development; and 3) exon assembly with ortholog search and alignment. The algorithm for exon capture bait development provides the user with a high flexibility in the adjustment of different bait parameters, including size, sequence similarity, nucleotide heterogeneity, and number of tiles. The final dataset produced by the program includes sequences of annotated and aligned exon orthologs without non-coding flanking regions. Currently, EXONtools has been approbated on Unix and Linux operating systems, using a dataset of 135 individuals of a non-model bony fish species and >15,000 exons captured per each individual. EXONtools is available in beta version by request.

Expanded summary*: The large number of species of indigenous freshwater fish fauna in tropical and subtropical oceanic archipelagos is represented by amphidromous gobies. Recent rapid agricultural expansion and urbanization processes as well as uncontrolled introductions of exotic invasive species on some tropical islands cause irreversible degradation of insular freshwater habitats and subsequent extinction of some native fish populations. Although many anthropogenic stressors that affect freshwater streams were well described in recent years, little is known about biology and population sustainability for the majority of threatened taxa. My dissertation project is aimed to investigate the population structure and larval recruitment patterns of native freshwater gobies from genus *Stenogobius* (Teleostei: Gobiidae) on Hawaiian, Marquesas and Society Islands based on comparison of their exomes, using the large number of individuals (>100). The experimental design of my research includes the following steps: whole transcriptome sequencing, *de novo* transcriptome assembly, transcriptome annotation, exon search, bait development, and exon capture sequencing. All these steps required implementation of many existing analytical approaches and sometimes new bioinformatics algorithms that are necessary for solving some specific tasks (e.g., isoform tests, exon search or bait design). "EXONtools" pipeline was developed in order to combine all these different programs and my personal codes into a single bioinformatics framework, written in Python programming language, that will allow all other researchers to perform similar studies based on exon capture sequencing in a much easier way. Exon capture sequencing is one of the most promising approaches in evolutionary and population genomics because it only obtains data from protein coding DNA regions (~2%), avoiding all other unnecessary sites. As a result, hundreds of individuals can be multiplexed and analyzed all together during a single high-throughput sequencing run, yet still providing a substantial amount of data required for deep phylogenomic reconstructions, large scale population studies, identification of candidate genes under selection, etc. Furthermore, in combination with *de novo* transcriptome sequencing (RNA-seq), the difficult task of capturing whole exomes for non-model organisms can also be accomplished. Currently, EXONtools has been approbated on Unix and Linux operating systems, using a dataset of 135 individuals of a non-model bony fish species and >15,000 exons captured per each individual.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-305

Are VBIRs like FlyClockbase the new Genome Projects?

Katherine Scheuer [1], Laurence Loewe [1,*]

[1]Laboratory of Genetics and Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, United States

Abstract: Genome projects have convincingly demonstrated that bundling similar tasks for batch processing can greatly increase the efficiency of biological research. Instead of a gene at a time, sequencing entire genomes provides a resource allowing biologists to efficiently query a genome. Current biology is hard to image without genomics. Why is it important to bundle similar tasks? (1) It drives tool development for faster task execution and reuse of setup overhead. (2) The resulting standardization enables better quality control. (3) It improves information flow as updates are best maintained by one team per genome, allowing other researchers to efficiently test hypotheses using the latest release. Bundling updates into versioned releases of a genome helps maintainers to integrate quality improvements and users to reference a precise state for reproducible results. Here we extend this idea to other areas of biological information by introducing Versioned Biological Information Resources (VBIRs). Each VBIR has its chosen topical focus that could be narrow or broad. VBIRs vary in size and could be implemented in different ways, but must provide a defined method for referencing and accessing past versions. Offering stable links to past variants of more causal VBIRs enables researchers to build more consequential VBIRs that could grow into complex networks of biological expertise. These could be used for simulating more complex biological systems, parameter inference, and hypothesis testing. We developed FlyClockbase, a VBIR focused on integrating observed time series of circadian clock core components from the fly *Drosophila melanogaster*. We used this VBIR to test the hypothesis that the variance of the peak times of the proteins PERIOD and TIMELESS differ significantly. Since variances are particularly sensitive to errors in the data, we conducted a human error analysis measuring relevant error rates in this VBIR. Our findings confirm results from the human error analysis literature and suggest that VBIRs benefit from investing in data quality and error control. Faulty input data can rarely be fixed automatically, but a compiler for VBIRs could detect errors, highlight ambiguities, and track their resolution by biological model curators. Before writing such a compiler we need a well-defined storage format. We started developing a radically open storage format that sits on top of standard file systems and is based on tab delimited tables of text organized in files within nested folders. This builds on the simplicity of widely used comma separated value files, and adds features that enable more expressivity in standardized ways. Our use of simple text tables enables experimental biologists to add new data using spreadsheet programs and computational biologists to programmatically access the data with minimal effort. It also facilitates much needed discussions between experimental biologists and computational biologists about data types for storing uncertain, imprecise, and contradictory biological information. We will need networks of hundreds or thousands of VBIRs like FlyClockbase in active use by their research communities, all stable enough for automating computational analyses and updates, in order to predict phenotypes from genotypes, personalize medicine, or simulate mechanistic fitness landscapes in evolutionary systems biology. More about FlyClockbase and VBIRs at https://doi.org/10.1101/099192

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-307

Clear: Composition of Likelihoods for Evolve And Resequence Experiments

Arya Iranmehr*, Ali Akbari, Christian Schloetterer, Vineet Bafna

Abstract:

The advent of next generation sequencing technologies has made whole-genome and whole-population sampling possible, even for eukaryotes with large genomes. With this development, experimental evolution studies can be designed to observe molecular evolution ``in-action" via Evolve-and-Resequence (E&R) experiments. Among other applications, E&R studies can be used to locate the genes and variants responsible for genetic adaptation. Existing literature on time-series data analysis often assumes large population size, accurate allele frequency estimates, and wide time spans. These assumptions do not hold in many E&R studies.

In this article, we propose a method--Composition of Likelihoods for Evolve-And-Resequence experiments (CLEAR)--to identify selection in short-term (as well as long-term), E&R experiments in sexual populations with small size. CLEAR takes whole-genome sequence of pool of individuals (pool-seq) as input, and properly addresses heterogeneous ascertainment bias resulting from uneven coverage. CLEAR also provides unbiased estimates of model parameters, including population size, selection strength and overdominance, while being computationally efficient. Extensive simulations show that CLEAR achieves higher power in detecting and localizing selection over a wide range of parameters, and is robust to variation of coverage. We applied CLEAR statistic to multiple E&R experiments, including, adaptation of Drosophila to a novel laboratory environment and a study of outcrossing Yeast populations, and identified multiple regions under selection with genomewide significance.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-306

Influence of alignment uncertainty on homology and phylogenetic modeling

Jia-Ming Chang [1,*], Cedric Notredame [2]

[1]Computer Science, National Chengchi University, Taipei, Taiwan, [2]Bioinformatics and Genomics, Centre for Genomic Regulation (CRG), Barcelona, Spain

Abstract: Most evolutionary analyses or structure modeling are based upon pre-estimated multiple sequence alignment (MSA) models. From a computational point of view, it is too complex to estimate a correct alignment. Hence, increasing or identifying signal inside sequence alignment has intensified over the last few years. During the presentation, I would like to share two approaches, homology extension and sampling, on this topic.

The first part, transmembrane proteins (TMPs) constitute about 20~30% of all protein coding genes. The relative lack of experimental structure has so far made it hard to develop specific alignment methods and the current state of the art (PRALINE™) only manages to recapitulate 50% of the positions in the reference alignments available from the BAliBASE2-ref7. We show how homology extension can be adapted and combined with a consistency based approach in order to significantly improve the multiple sequence alignment of alpha-helical TMPs. TM-Coffee is a special mode of PSI-Coffee able to efficiently align TMPs, while using a reduced reference database for homology extension. Our benchmarking on BAliBASE2-ref7 alpha-helical TMPs shows a significant improvement over the most accurate methods such as MSAProbs, Kalign, PROMALS, MAFFT, ProbCons and PRALINE™. We also estimated the influence of the database used for homology extension and show that highly non-redundant UniRef databases can be used to obtain similar results at a significantly reduced computational cost over full protein databases.

The second part, homology and evolutionary modeling are the most common applications of MSAs. Both are known to be sensitive to the underlying MSA accuracy. In this work, we show how this problem can be partly overcome using the transitive consistency score (TCS), an extended version of the T-Coffee scoring scheme. Using this local evaluation function, we show that one can identify the most reliable portions of an MSA, as judged from BAliBASE and PREFAB structure-based reference alignments. We also show how this measure can be used to improve phylogenetic tree reconstruction using both an established simulated data set and a novel empirical yeast data set. For this purpose, we describe a novel lossless alternative to site filtering that involves overweighting the trustworthy columns. Our approach relies on the T-Coffee framework; it uses libraries of pairwise alignments to evaluate any third party MSA. Pairwise projections can be produced using fast or slow methods, thus allowing a trade-off between speed and accuracy. We compared TCS with Heads-or-Tails, GUIDANCE, Gblocks, and trimAl and found it to lead to significantly better estimates of structural accuracy and more accurate phylogenetic trees.

References:

• PSI/TM-Coffee: a web server for fast and accurate multiple sequence alignments of regular and transmembrane proteins using homology extension on reduced databases. *Nucleic acids research* 44, W339–343(2016).
• TCS: a web server for multiple sequence alignment evaluation and phylogenetic reconstruction. *Nucleic acids research* 43, W3–6 (2015).
• TCS: a new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction.*Molecular biology and evolution* 31, 1625–37 (2014).
• Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. *Bmc Bioinformatics* 13, S1 (2012).

Website:

• PSI/TM-Coffee http://tcoffee.crg.cat/tmcoffee
• TCS http://tcoffee.crg.cat/tcs

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-304

Measuring the amount and distribution of positive Darwin selection using both polymorphism and divergence in Drosophila

Hao Yang [1,*], Ao Lan [1], Qipian Chen [1], Suhua Shi [1], Chung-I Wu [1 2]

[1]School of Life Sciences, Sun Yat-Sen University, Guangzhou, China, [2]Department of Ecology and Evolution, University of Chicago, Chicago, United States

Abstract: Determine the amount and the distribution of positive Darwin selection among genome has long been one of the central goals in the evolutionary studies. Benefit from the recent development in sequencing technology, in this study, 5 *Drosophila* species and over 300 individuals of *Drosophila melanogaster* are used for McDonald-Kreitman test (MK test) as well as the method of Phylogenetic Analysis by Maximum Likelihood (PAML). Interestingly, on a gene by gene basis, these two methods have very poor overlaps (267 positive selected genes by MK test and 100 by PAML, with only 11 shared). However, when comparing by using other parameters such as the number of positive selected sites detected, it is clear that both of the methods provide rather similar trends. Based on these findings, we suggest that during the evolutionary history, positive Darwin selection could be both weak and wide-spread across the genomes. Thus, MK test and PAML, with their distinct assumptions behind, pick out very different sets of genes by using very different angles.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

OTH-CR1

Engineering a real-time HIV surveillance platform

Simon Frost [1,*], Mukarram Hossain [1], Art Poon [2], Sergei Kosakovsky Pond [3], Steven Weaver [3], Erik Volz [4], Trevor Bedford [5], Richard Neher [6], Sten Vermund [7 8], Timothy Sterling [9], Ann Dennis [10], Marcia Kalish [8]

[1]Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom, [2]Pathology and Laboratory Medicine, Western University, London, Canada, [3]Biology, Temple University, Philadelphia, United States, [4]Infectious Disease Epidemiology, Imperial College London, London, United Kingdom, [5]Vaccine and Infectious Disease, Fred Hutchinson Cancer Research Center, Seattle, United States, [6]Biozentrum, University of Basel, Basel, Switzerland, [7]School of Public Health, Yale University, Newhaven, [8]Institute for Global Health, [9]Medicine, Vanderbilt University, Nashville, [10]Medicine, UNC Chapel Hill, Chapel Hill, United States

Abstract: Background

Recent outbreaks of HIV among black men who have sex with men (MSM) in Jackson, Mississippi and among people who inject drugs in Scott County, Indiana demonstrate the potential for dramatic, rapid spread of HIV in some settings, which can be informed by phylogenetic analysis of viral sequence data. Analysis of HIV sequences is usually retrospective, involves the use of multiple, often web-based tools. We developed a Python software package, nextHIV, an open-source system for real-time analysis of HIV sequence data, which is self-contained, fast, and portable.

Methods

For data privacy, analyses such as viral subtyping, drug resistance interpretation are implemented locally, rather than using web-based tools. To allow analysis of sequence data in near real-time, we used algorithms that are fast, can be parallelized, and can be run in an 'on-line' setting. These include codon-informed sequence alignment; nucleotide distance calculation; machine-learning approaches for viral subtyping; and sequential approaches to phylogeny reconstruction. Reports are used to visualize the data, which are generated as static HTML files, to allow interaction with the data that is separated from the database, improving data security. To permit remote deployment on different operating systems, and to ensure that the software runs identically on all platforms, the package and all dependencies can be installed as a Docker container.

Results

To simulate surveillance in real time, we applied our platform to retrospective data from the Vanderbilt Comprehensive Care Clinic, comprising 4,728 sequences from 2916 individuals sampled between 1998 and 2015 from middle Tennessee, US. Our analysis identified active transmission of HIV, particularly among young men who have sex with men, as well as clusters of non-B subtypes consistent with local transmission.

Conclusions

We have developed a self-contained system for automated molecular epidemiological analysis, which has the potential to inform public health strategies for reducing transmission. While targeted at HIV, it is also portable to the surveillance of other viral infections.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-303

Three-dimensional Growth and Migration of Cancer Cell Population

Li Guanghao [1,*], Yang Zuyu [1], Lu XueMei [1], Wu Chung-I [1,1]

[1]Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Abstract: Growth models are widely used to uncover the evolution process of tumor, which is difficult to be traced practically. Studies related to tumor growth models started very early and various models have been proposed including exponential growth, surface growth and so on. These models and tumor genome sequencing results indicate that genetic diversity in tumor is extremely high. In order to figure out the dynamic process of tumor population, 3D growth models are required to construct. At the same time, the three-dimensional sampling for a tumor case is carried out in high density. Subsequently, we sequence the sample in whole genome level to obtain genetic variation in high resolution. With these mutations, we can infer evolution history and evolutionary forces of tumor cell population. Combining the inference result with the constructed growth model, tumor three-dimensional growth and migration can be estimated. We simulated tumor growth with somatic cells growing in Wright-Fisher model and mutations occurring in infinite-site model. The result shows that spatial distribution of tumor subclones presents a radial pattern and the diversity of subclones is decided by ancestor cell.
Note: ABSTRACT EXTENSION

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-309

Phylogenetic accuracy of permutated samplings (PAPS) pipeline: taxon sampling influence on stability and speciation rates in Terrabacteria

Ashley Superson [1,*], Douglas Phelan [1], Michael Culver [1], Anna M. Spagnuolo [1], Fabia U. Battistuzzi [1]

[1]Oakland University, Rochester, United States

Abstract: Tree of Life (TOL) reconstruction efforts are continually improved by sequencing advancement but also hindered by various biological and analytical factors that result in conflicting phylogenetic signals. In any phylogeny, two aspects are dominant: topology (ancestor-descendant relationship) and patterns of speciation rates (clustering of nodes through time). While for the latter there is only sparse information, for the former current TOL reconstruction studies have exposed conflicting phylogenies for the Terrabacteria superphylum, particularly in the placement of the Deinococcus-Thermus (DT) phylum. We hypothesize that at least part of this uncertainty could be caused by the DT species representation that is much lower compared to that of other phyla. To test this hypothesis, we constructed the PAPS pipeline that allows to easily build permutations on an aligned dataset to determine how altered sampling scenarios affect phyletic signal accuracy and speciation rates. We used a discrete Robinson-Foulds metric to quantify the level of discord among permutated maximum-likelihood trees while for rates we utilized ordinary differential equations to model speciation events in timetrees. A resolved TOL and the placement of specific speciation events within the tree are of crucial importance to reach a greater understanding of the origin of life, the evolutionary trajectory of major species, and the resulting adaptations life has made in approximately four billions of years of evolution. Our data shows that different taxon samplings affect phylogenetic reconstruction suggesting that to obtain a stable and accurate TOL sequencing efforts should be more evenly distributed across taxonomic categories.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-310

Standard mapping protocols misestimate sex-linked gene expression

Kimberly Olney [1,*], Sarah Brotman [1], Melissa Wilson Sayres [1 2]

[1]School of Life Sciences, [2]Center of Evolutionary Medicine , Arizona State University , Tempe, United States

Abstract: There are several challenges to accurately inferring levels of transcription using RNA-sequencing (RNAseq) data, including detecting and correcting for reference genome alignment bias. However, one potential confounder of RNAseq analysis can result when applying a standardized pipeline to samples of different sexes in species with chromosomal sex determination. Because of the homology between the human X and Y chromosome, we expect that mis-mapping will routinely occur between these two chromosomes, artificially affecting estimates of sex-linked gene transcription. For this reason we tested alternative alignment scenarios on RNAseq samples from the brains of 5 genetic female and 5 genetic males to assess how inferences of differential gene expression patterns change depending on the reference genome. We first applied a standard alignment protocol where we aligned all individuals to the entire human reference genome, and computed differential expression between the set of male and female samples. Then, we realigned the female samples (46,XX) to the human reference genome with the Y chromosome removed, and the male samples (46, XY) to the human reference genome (including the Y chromosome), but with the pseudoautosomal regions of the Y chromosome masked out. The new strategy called 33 additional genes as being differentially expressed between the two sexes when the genetic female samples were mapped to the reference without the Y chromosome and the genetic males were mapped to the reference without the Y PARs. This research is providing essential insight into correcting genome alignment bias that can be used in future studies.

Expanded summary*: Standard mapping protocols misestimate sex-linked gene expression

Sarah M. Brotman, **Kimberly C. Onley**, and Melissa A. Wilson Sayres

Background: Genetic males and females share highly similar genomes, only differing in the sex chromosomes, yet males and females are morphologically and physiologically distinct. Sexual dimorphism in aspects of human biology, such as development, physiology, metabolism, susceptibility to disease, and wound healing, develop through sex-specific gene regulation. Interestingly, the amount of the genome that is differentially expressed can be substantial, even in non-reproductive tissues. Characterizing biases in gene expression is central to understanding sex-specific physiology and sexual dimorphism. However, identifying sex-biased expression is complicated by lack of a standard mapping protocol for sex-linked genes and the challenges of gametologous sex-linked sequences. The X and Y chromosomes were once homologous autosomes that could undergo recombination, however due to recombination suppression on the Y chromosome, they are no longer able to recombine along the entire length with the exception of the pseudoautosomal regions (PARs). Because of the homology between the human X and Y chromosome, we expect that mis-mapping will routinely occur between these two chromosomes, artificially affecting estimates of sex-linked gene transcription. For this reason we tested alternative alignment scenarios on RNAseq samples from brain, blood, heart, liver, and kidney tissues in five genetic males (46. XY) and five genetic female (46, XX). We show how inferences of differential gene expression patterns change depending on the reference genome.

Reference mapping bias: The GENCODE reference genome was obtained in addition to our creation of two equivalent reference genomes: one without the Y chromosome, and one with the PARs masked in the Y chromosome. We hypothesized that there would be reference mapping bias when we aligned genetic females (46,XX) to the reference with the Y chromosome versus when we mapped the samples to the reference without the Y chromosome. STAR read aligner was used to map transcript reads to each of the reference genomes, which was followed by running CuffDiff to determine the differential expression between samples mapped to the original reference genome versus the modified reference genomes. The preliminary results from the brain samples indicate that there are 33 more genes being called as differentially expressed when the genetic female samples were mapped to the reference without the Y chromosome and the genetic males were mapped to the reference without the Y PARs. Of the 33 additionally called genes that were differentially expressed, some were mapped to regions in the X chromosome, Y chromosome, as well as in some autosomes, indicating that more than just the pseudoautosomal region was mis-mapping when aligning both genetic males and genetic females to the standard reference genome that includes the X and Y chromosome. The implication of reference mapping bias puts into question

the common practice of using of one standard reference genome for both sexes in a study. Our goal is to make our reference genomes and pipeline publically available on GitHub so that researchers can use it in their studies to limit reference mapping bias,

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-312

Re-evaluating the target of selection within FOXP2 suggests functional divergence among diverse human populations

Elizabeth Atkinson [1,*], Gregory Smith [1], Amanda Audesse [2], Ashley Webb [3], Sohini Ramachandran [4], Brenna Henn [1]

[1]Ecology and Evolution, Stony Brook University, Stony Brook, [2]Neuroscience, [3]Molecular Biology, Cell Biology and Biochemistry, [4]Ecology and Evolutionary Biology and Center for Computational Molecular Biology, Brown University, Providence, United States

Abstract: *FOXP2* was identified due to its critical role in human speech. Arguments for a recent selective sweep in *Homo sapiens* were thrown into doubt by the conflicting timeline that emerged from aDNA evidence. Here, we comprehensively reanalyze *FOXP2* in hundreds of next-generation genomes and exomes from globally distributed human populations. Intriguingly, demographically-controlled Tajima's *D* calculations do not support a recent selective sweep in *FOXP2*. Instead, these and other metrics, including haplotype networks and ARGs, suggest balancing selection in African populations. Specifically, we identify three major, common haplotypes segregating in modern humans that span a narrow region in *FOXP2* intron 9 that is significantly ($P<0.001$) evolutionarily conserved across vertebrates. This region harbors high-GERP SNPs that are derived in humans compared with all other primates, including archaic hominins, and is a statistically significant genomic outlier in these datasets. Strong evolutionary constraint among taxa but variability within *Homo sapiens* is compatible with this locus having a major functional role unique to humans. We examine if these haplotypes affect the function of the FOXP2 transcription factor through the production of alternative coding isoforms or differential expression of FOXP2/target genes. Using RT-PCR, we observe expression of a potentially novel exon in the region of interest in some RNA transcripts from human immortalized brain cells (U87 cells) despite being in an annotated intron. We additionally test for the presence and alteration of the relative expression of FOXP2 isoforms and targets with relation to the region of interest after modifying significant SNPs using CRISPR techniques.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

OTH-CR3

Comparative genomics of Plasmodium parasites using the web-based CoGe platform

Andreina Castillo Siri [1,*], Eric Lyons [1]

[1]University of Arizona, Tucson, United States

Abstract: CoGe is a publicly available online platform that provides a suite of tools for the analysis, visualization, and storage of genomic data. CoGe currently has over 31,000 genomes from 17,000 organisms and contains tools to let research quickly integrate and analyze new genomic data. Eukaryote parasites have reduced genomes, lost many metabolic functions, and developed specialized genomic components. Here, we demonstrate CoGe's capabilities for comparative genomic analysis of several Plasmodium species. We focused on Plasmodium due to their medical relevance, widespread distribution, and unique evolutionary history. Fully sequenced and annotated Plasmodium genomes, as well as near-complete genome assemblies, are included to show the versatility of CoGe's capabilities. The example analyses we present include:

- A characterization of general trends of GC content variation and amino acid usage across the genus.
- An exploration of the origin and location of structural genomic evolutionary events (inversion, deletions, and fission/fusion events) that changed genome architecture.
- The identification of variations in the strength and direction of natural selection across the length of genomes.
- An examination of the evolutionary dynamics of specific genetic elements via the analysis of syntenic regions.
- Unraveling the evolutionary history of highly variable multigene families at the intraspecific level.

The web-based CoGe platform is a highly efficient system in the study of Plasmodium genomes and other eukaryotic parasites. CoGe is freely available online at https://genomevolution.org/coge/. An in-depth tutorial of the analyses presented here is available at https://genomevolution.org/wiki/index.php/Using_CoGe_for_the_analysis_of_Plasmodium_spp

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-316

Using genealogical properties to identify selective sweeps in the human thousand genomes

Martina Rauscher [1], Johannes Wirtz [1], Thomas Wiehe [1,*]

[1]University of Cologne, Institute for Genetics, Cologne, Germany

Abstract: There is still disagreement about mode, strength and rate of selective sweeps in human. Identifying loci which underwent recent selective sweeps is difficult because the traces are typically obscured by other evolutionary and demographic forces, such as genetic drift in small populations or population substructure. To detect candidate loci of selective sweeps we take here an approach which considers genealogical relationships among individuals and the topological properties of the inferred coalescent tree.
Since selective sweeps can produce highly unbalanced coalescent tree topologies in regions close to a selective sweep site, we have devised a statistic, called T3, to detect bias in tree balance. Under neutral evolution and panmixis, T3 is approximately standard-normally distributed, while selective sweep genealogies lead to negatively biased T3 values. Since this property is inherited also to subsamples, iterated sub-sampling and re-testing provides a means to increase power of the T3 test which is different from lowering the significance level.
We present the results of genome wide screens of the T3 test with subsampling applied to the 26 populations of the final release of the human 1000 genomes project.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-313

clade-based annotations in ensembl

Konstantinos Billis [1,*], Carlos García Girón [1], Leanne Haggerty [1], Thibaut Hourlier [1], Osagie Izuogu [1], Daniel N. Murphy [1], Rishi Nag [1], Fergal J. Martin [1], Paul Flicek [1], Bronwen Aken [1]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, United Kingdom

Abstract: Numerous large-scale genome sequencing initiatives are underway. The creation of a large number of vertebrate genome assemblies will lead to a valuable resource for the life sciences. Further genomic analysis, such as consistently annotating the genes within these assemblies, will be crucial to enable meaningful studies of the species they represent.
In response, we have rebuilt the Ensembl gene annotation system, using our eHive pipeline management framework. Our approach is based on a combination of RNA-seq alignments, annotation projection via whole genome alignments and protein-to-genome alignments using selected UniProt proteins. These new pipelines reduce the gene calling process from months to days, while continuing to provide the high-quality gene annotations associated with Ensembl.
We are currently using our new system to annotate rodents and primates, with more clades planned in the near future. Annotating in a clade-based manner will ensure maximal consistency and efficiency in terms of gene annotations and comparative genomics analysis.
The annotated assemblies of several rodent and primate species will be available in Ensembl later this year, in addition to updated gene sets for several of our key species. Where available, annotated genomes will include RNA-seq data, which can be viewed on the Ensembl genome browser. All data will be available via our FTP site, REST API or Perl API.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

OTH-CR4

Reproducible Reconstruction, Analysis and Visualization of Phylogenomic Data using ETE.

Jaime Huerta-Cepas*

Abstract:

The Environment for Tree Exploration (ETE) is a computational framework that simplifies the reconstruction, analysis, and visualization of phylogenetic trees and multiple sequence alignments. ETE v3 [1] features numerous improvements in the underlying library of methods, and provides a novel set of standalone tools to perform common tasks in comparative genomics and phylogenetics. Those features include (i) reproducible reconstruction of gene-based and supermatrix-based phylogenies using a single command, (ii) testing and visualizing evolutionary models, (iii) calculating distances between trees of different size or including duplications, and (iv) providing seamless integration with the NCBI taxonomy database.

Most notably, the *ete-build* tool provides a unified interface to wrap the execution of reproducible phylogenetic workflows, comprising the reconstruction of gene-trees and supermatrix-based species trees. To do so, ETE relies on a versioned collection of external tools that are transparently installed and executed upon request. A single command is used to configure and launch complex phylogenetic pipelines, covering sequence alignment, trimming, substitution-model testing, tree inference, and image rendering. In addition, the supermatrix-based reconstruction mode permits to build and concatenate multiple sequence alignments with ease, simplifying the inference of species trees based on multiple genes. Advanced options allow to automatically switch from amino-acid to nucleotide alignments based on sequence identity, resuming the execution of workflows, or even testing multiple strategies in parallel. As an example, a single command line can be used to test several alignment methodologies or phylogenetic inference programs simultaneously, making the tool particularly suitable to run phylogenomic pipelines.

Recent applications of ETE include the computation of over one million phylogenetic trees for the EggNOG v4.5 database [2] and the automated prediction of orthology at the genomic scale for functional annotation purposes [3]. ETE is freely available at http://etetoolkit.org

[1] Huerta-Cepas, Jaime, François Serra, and Peer Bork. 2016. “ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data.” *Molecular Biology and Evolution* 33 (6): 1635–38.

[2] Huerta-Cepas, Jaime, Damian Szklarczyk, Kristoffer Forslund, Helen Cook, Davide Heller, Mathias C. Walter, Thomas Rattei, et al. 2016. “eggNOG 4.5: A Hierarchical Orthology Framework with Improved Functional Annotations for Eukaryotic, Prokaryotic and Viral Sequences.” *Nucleic Acids Research* 44 (D1): D286–93.

[3] Huerta-Cepas, Jaime, Kristoffer Forslund, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork. 2016. “Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper.” *bioRxiv*. doi:10.1101/076331.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-314

Phylogenetic inference is affected by closure, and non-closure, of substitution models.

Jeremy Sumner [1], Michael Charleston [1,*], Michael Woodhams [1], Barbara Holland [1]

[1]University of Tasmania, Hobart, Australia

Abstract: In phylogenetics, time reversible (TR) DNA mutation models have been popular for decades. A more recent development is the hierarchy of Lie-Markov (LM) models, whose defining property is that their Markov matrices are closed under multiplication. This closure property means that two consecutive LM processes can be represented as another LM process (in the same model) – but most TR models don't have this property; so for example a GTR process followed another GTR process need not be a GTR process overall.We have chosen a set of LM models from the many permitted by this Lie-Markov property, constructed using sensible symmetries arising from biology. In particular, we allow transitions to behave differently from transversions, but insist all transitions be on an equal mathematical footing with each other, and similarly all transversions. We have investigated how the closure property, or lack thereof, affects our ability to accurately recover phylogeny, when the substitution process is permitted to change through time. We generated sequence data on trees under a range of substitution models both with and without multiplicative closure, where we allowed different instances of the same model to occur on consecutive branches. We used both LM and TR models to estimate the underlying tree for all simulations.We found that, when we have heterogeneous DNA mutation processes, but treat them as homogeneous when inferring the tree, then Lie-Markov models are generally superior to time reversible models in reconstructing the phylogeny.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Computation and reproducibility in molecular evolution

POB-315

Evolveathon I: A competition to detect selective sweeps from genome scale polymorphism data where targets can be independently validated.

Daniel Jeffares [1,*]

[1]Biology, University of York, York, United Kingdom

Abstract: **Summary:** Genome scale data polymorphism can be used to detect selective sweeps that have occurred in the recent past. But without any 'positive control set' there will always be doubt about the efficacy of methods. I describe a proposal to generate such a positive control data set, by evolving yeast populations in laboratory culture until sweeps occur (detected by pooled sequencing, not released to contestants) and sequencing haplotypes (which are released to contestants). **Rationale:** Simplistic sweeps can be modelled computationally, but many aspects of genomes, populations and the genetic architecture of traits are not sufficiently well known to simulate. For example, given the polygenic and epistatic genetic architecture of quantitative traits, polygenic soft selective sweeps may be the norm. Sweeps from standing variation, influenced by elements of population structure are also likely. The result is that the detection of recent adaptive evolution from real data, while supported models, is frequently carried out without a convincing sense of the sensitivity or accuracy of the method. **"Evolveathon I" competition proposal:** To assist the development of methods to detect selective sweeps, I propose to run a competition where contestants detect selective sweeps. I suggest Evolveathon should proceed as follows.

We will:
1. Culture populations of haploid fission yeast strains in the laboratory. Such populations always show large changes in allele frequencies due to selection (preliminary data confirms this). Cultures will undergone cycles of mitotic growth interspersed with meiotic recombination. **2.** Sequence population pools at various time points to confirm genome-wide changes in allele frequency (which will detect sweeps at later time points with high sensitively and accuracy). **3.** Sequence individual strains (haplotypes) from pools at various time points where changes in allele frequency are detectable to different degrees. Strains will be made available at the end of the competition. **4.** Release SNP and sequence data (from strains, not from pools), staging releases so the task becomes progressively simpler as allele frequency changes become greater.
5. Assess the accuracy of any submissions and report back at future SMBE meetings, via publications etc.
6. Follow up features of the sweeps such as time scale, trait details and selection coefficient (if funds allow).

Interested parties are encouraged to: Get in touch. Download data and submit the predictions for sweep locations, along with any parameters they choose to estimate.

Details: Populations will be mixtures of at least 57 Schizosaccharomyces pombe (fission yeast) strains. We will culture these until allele frequency changes are clear. Strains have all been sequenced with analysis of population structure with extensive analysis of quantative traits (1). Strains show strong intrinsic reproductive isolation, so will retain some 'population structure' in vitro. Structural variants are characterised and known to influence traits and reproductive isolation2. Sequences will be at least 50x (probably greater) for pools. We will sequence at least 40 strains for each time point, to at least 5x (probably greater).

I welcome expressions of interest in the competition, suggestions for modifications of the methods and any other comments. Contact: daniel.jeffares@york.ac.uk

References:
1. Jeffares et al. Nature Genetics 47, 235–241 (2015).
2. Jeffares et al. Nat Commun 8, 14061 (2017).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-159

Sexual selection gives rise to non-parallel evolution

Ana Marija Jaksic [1 2,*], Neda Barghi [1], François Mallard [1], Viola Nolte [1], Kathrin Otte [1], Christian Schlötterer [1]

[1]Institut für Populationsgenetik, Vetmeduni Vienna, [2]Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Vienna, Austria

Abstract: Sexual selection represents one of the strongest directional evolutionary forces, with a diverse set of traits being affected. In *Drosophila* the high rate of sequence evolution, which was observed for accessory gland proteins (ACPs) and seminal fluid proteins (SFPs), has been attributed to sexual selection. With such a large number of potential target genes for sexual selection, it is not clear how repeatable the response to sexual selection is. We addressed this question by performing a highly replicated experimental evolution study in *D. simulans*. Starting from the same founder, 10 populations independently evolved for 100 generations. For each population, RNA-Seq was performed in three replicates to identify genes with differential expression between the 10 independently evolved populations. We classified 924 expressed reproductive genes as potential targets of sexual selection and found 40% (368 genes) to be differentially expressed between the evolved populations, which is significantly higher than expected by chance ($p=1.5x10^{-5}$). Consistent with the previously reported evidence for sexual selection, many genes directly involved in sperm competition (66%), including ACPs (50%) and SFPs (66%) were among the differentially expressed genes. We conclude that the large number of genes that can be targeted by sexual selection resembles a quantitative trait and does not exhibit patterns of parallel evolution at the gene expression level. We will discuss to what extent sexual selection is mediated by gene expression differences rather than amino acid replacements.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-157

I wanna be like you: The prevalence of molecular convergence in closely related organisms

Lucy Holloway [1,*], Martin Hughes [2], Mark Wilkinson [2], Davide Pisani [1]

[1]Palaeobiology, University of Bristol, Bristol, [2]Natural History Museum, London, United Kingdom

Abstract: The closest thing we currently have to 'laws of convergence' is a generally accepted assumption that homoplasy is less prevalent in molecular than morphological data, and that homoplasy is more common among organisms that are closely related. This study aims to test and formalise these concepts by comparing molecular data from organisms across the tree of life. Data types tested include nucleotide and amino acid sequences as well as protein families and presence-absence data for protein domains. Previous work using the same method of analysis on morphological data has demonstrated that morphological convergence is more likely to occur among more closely related organisms. Preliminary results from this study suggest that not only is homoplasy prevalent in molecular data, similar patterns of convergence can be seen in molecular as in morphological data. In particular, homoplasy of protein domains within Metazoa is more likely to be observed among more closely related taxa.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-161

Sex-limited polymorphisms through repeated but unique changes to pdm3

Emily Delaney [1,*], Artyom Kopp [1]

[1]Evolution & Ecology, University of California-Davis, Davis, United States

Abstract: Sexually dimorphic traits are ultimately the product of sexually dimorphic gene expression. Although the molecular changes responsible for transitions between monomorphic and dimorphic gene expression have been identified for a few genes in a handful of species, it remains unclear whether sexually dimorphic gene expression generally evolves via parallel mutations. To determine what types of genetic changes are involved in the origin of sexually dimorphic alleles (e.g. gaining or losing of activating or silencing *cis*-regulatory elements), we analyzed the genetic basis of a convergent female-limited color polymorphism in six species from the *Drosophila montium* subgroup (a clade within the *melanogaster* species group). In each species, females have dimorphic light or dark abdominal pigmentation whereas males have only monomorphic pigmentation (light or dark, but always a single color). We repeatedly mapped dimorphic female pigmentation to a locus containing the gene *pdm3*, but to different intronic and intergenic regions, suggesting independent changes to *pdm3*. We are now resequencing this locus in wild and admixed populations to identify the precise nucleotides associated with female pigmentation, and are testing how these nucleotides alter *pdm3* expression using transgenic reporter assays and immunohistochemistry. In addition to comparing female-limited alleles among species with female-limited color polymorphisms, we also compare them to orthologous alleles from sexually monomorphic species in the *montium* subgroup and to *D. melanogaster*, a close relative that lacks female-limited color polymorphism. Our extensive comparative genetic and developmental analysis of a convergent set of sex-limited alleles will reveal whether parallel genetic changes are necessary for sexually dimorphic expression.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

OW-CE3

All roads lead to adaptation: Genomic signatures of parallel evolution in Drosophila simulans

Neda Barghi [1,*], Raymond Tobler [1 2], Viola Nolte [1], Christian Schlötterer [1]

[1]Institut für Populationsgenetik, Vetmeduni, Vienna, Austria, [2]Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Adelaide, Australia

Abstract: Highly replicated experiments with microbes indicated that replaying the tape of life results in more parallel evolution than previously anticipated. This high reproducibility is attributed to genetic constraints, resulting in different mutations in the same genes or pathways. Little is known about the reproducibility of adaptation from standing variation in sexual eukaryotes, where starting frequency, effective population size, and reservoir of mutations come into play. Furthermore, the adaptive traits are likely to be more complex than for microorganisms. We used the most highly replicated E&R study in *Drosophila* to study parallel evolution. The analysis of 10 replicate *Drosophila simulans* populations adapting to a hot environment uncovered a large number of candidate SNPs. To account for linkage, these SNPs were clustered into haplotype groups, each defining a selected genomic region. Interestingly, few clusters were selected in all replicates and most of them increased in frequency only in 1-5 replicates. Consistent with genetic drift being a major factor determining the heterogeneity among replicates, we found a strong positive correlation between the starting frequency of the selected clusters and the number of replicates showing a selection signature. Nevertheless, the clusters were significantly enriched for genes involved in oxidative phosphorylation pathway and TCA cycle, as well as genes with monosaccharide transmembrane transporter, triglyceride lipase, and endopeptidase activities. Moreover, all evolved replicates had a higher fecundity and metabolic rate than their ancestral populations. We conclude that natural *D. simulans* populations harbor sufficient standing variation to fuel parallel evolution of fitness and metabolism despite low parallelism on the gene level.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

OW-CE5

Ancient convergent loss of a metabolic gene's function yields deleterious consequences for modern marine mammals

Wynn K Meyer [1,*], Robert K Bonde [2], Maria Chikina [1], Clement E Furlong [3 4], Joseph C Gaspard III [5], Jerrica Jamison [6], Rebecca J Richter [3 4], Nathan L Clark [1]

[1]Department of Computational and Systems Biology, University Of Pittsburgh, Pittsburgh, PA, [2]Wetland and Aquatic Research Center, U.S. Geological Survey, Gainesville, FL, [3]Department of Genome Sciences, [4]Department of Medicine, University of Washington, Seattle, WA, [5]Pittsburgh Zoo and PPG Aquarium, [6]Dietrich School of Arts and Sciences, University Of Pittsburgh, Pittsburgh, PA, United States

Abstract: Convergent evolution, the independent acquisition of similar traits across multiple evolutionary lineages, provides a means of investigating the genes and phenotypes underlying adaptation. In one striking example, three mammalian lineages (cetaceans, pinnipeds, and sirens) independently adapted to marine lifestyles, with accompanying changes in phenotypes related to diet, morphology, and respiration. To characterize molecular changes associated with, and potentially underlying, these phenotypic changes, we identified genes with unique patterns of evolution specifically on marine branches of the mammalian phylogeny. One such gene had been pseudogenized only in marine lineages: PON1, a HDL-associated protein that combats lipid oxidation in the bloodstream. We determined that the loss of PON1 was ancient, predating splits among cetaceans and sirens. To assess whether shifts in diet, cholesterol levels, or oxidative environment may have contributed to PON1's loss, we investigated evolutionary patterns at gene sets involved in these processes. In addition to its metabolic role, PON1 is also the primary line of defense against the neurotoxic effects of organophosphate pesticides. Its recent loss therefore suggests that manatees may experience negative consequences in the presence of agricultural runoff and increasing pesticide use. We confirmed PON1's pseudogene status in a sample of seven Florida manatees and the dugong and further ruled out compensatory mechanisms in the bloodstream by analyzing the catalytic capacity of manatee plasma against three organophosphate compounds. We thus demonstrate that an ancient convergent loss of PON1 in marine mammals, likely due to changes in lipid metabolism, has resulted in increased risk for organophosphate-induced neurotoxicity in manatees.

Expanded summary*: Summary

This project uses evolutionary analyses to investigate the causes of a striking pattern of loss across marine mammals of a gene relevant for metabolism and protection from pesticides: Paraoxonase 1 (PON1). We use molecular and biochemical assays to predict the consequences of PON1 loss in the manatee, a species at risk of exposure to organophosphate pesticides due to its habitat of shallow coastal waters in proximity to agricultural runoff and targeted insecticide use.

We initially identified PON1 as a gene of interest through a study of convergent evolution in marine mammals [1], which demonstrated that hundreds of genes experienced shifts in evolutionary rates on mammalian lineages associated with marine environments. To determine whether some of these rate shifts may have been associated with loss of gene function, we searched for frameshifts and premature stop codons in all coding sequences included in the analysis. We then used the software BayesTraits [2,3] to test for an association of the presence or absence of gene function with species' marine status, correcting for phylogeny. We identified over 100 significantly associated genes, and an enrichment of genes with functions related to lipid metabolism. This category includes PON1, the only gene lost in all three marine lineages and no others.

Based on previous functional work on PON1 (e.g., [4]), as well as its strong evolutionary rate covariation [5] with genes involved in beta-oxidation of fatty acids, we formed three hypotheses about which of its functions may have led to its loss in marine species: selective effects related to changes in fatty acid composition of a marine diet, changes in HDL and/or LDL cholesterol levels in the bloodstream, and changes in oxidative environment due to submersion. We investigated these by choosing gene sets that might be similarly affected by each of these selective shifts and determining whether their evolutionary rates shifted in association with the loss of PON1.

In addition to investigating potential factors surrounding PON1's initial loss, we also performed follow-up analyses to predict the consequences of this loss in manatees, the species at highest risk of organophosphate exposure. We performed Sanger sequencing to confirm the predicted frameshift and premature stop codons in PON1 in a sample of seven Florida manatees, and we additionally sequenced two dugongs, enabling us to infer when PON1 was lost in sirens. To determine whether this gene loss led to uncompensated loss of organophosphate catalytic ability in the bloodstream, we tested biochemical activity of manatee plasma against chlorpyrifos, paraoxon, and diazoxon. We found that this activity was substantially lower than that of almost all mammalian plasma tested to date, leaving manatees potentially highly susceptible to organophosphate poisoning.

Significance statement

This work has broad-reaching implications for both the study of convergent evolution and the conservation of manatees. The enrichment of lipid metabolism genes in the set with significant associations between gene loss and the marine environment implies that investigating how marine species acquire, process, and store lipids will be critical in understanding their adaptation to a novel environment. PON1 provides a key example of molecular changes whose fitness consequences depend upon their environment: ancestral marine mammals experienced a drastic reduction in constraint at PON1, enabling its loss, yet today this loss may put manatees at risk due to anthropogenic changes to their environment. These predicted consequences for manatees should inform future conservation efforts and recommendations regarding acceptable levels of organophosphate use near manatee habitats.

References

1. Chikina M, Robinson JD, Clark NL (2016) Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. Mol Biol Evol 33: 2182–2192. doi:10.1093/molbev/msw112.
2. Pagel M, Meade A (2006) Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible- Jump Markov Chain Monte Carlo. Am Nat 167: 808–825. doi:10.1086/503444.
3. Pagel M, Meade A (2013) BayesTraits.
4. Furlong CE, Li W-F, Shih DM, Lusis AJ, Richter RJ, et al. (2002) Genetic Factors in Susceptibility: Serum PON1 Variation Between Individuals and Species. Hum Ecol Risk Assess An Int J 8: 31–43. doi:10.1080/20028091056700.
5. Clark NL, Aquadro CF (2010) A novel method to detect proteins evolving at correlated rates: identifying new functional relationships between coevolving proteins. Mol Biol Evol 27: 1152–1161. doi:10.1093/molbev/msp324.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-152

Peto's Paradox and Beyond: The Evolution of Cancer Suppression across Mammals

Marc Tollis*, Lisa Abegglen [1], Aleah Caulin [2], Andrea Cabrerra [3], Andrew Webb [4], Nader Pourmand [5], Richard Green [5], Jooke Robbins [6], Mary O'Connell [7], Per Palsbøll [3], Joshua Schiffman [1], Carlo Maley [8]

[1]University of Utah School of Medicine, Salt Lake City, [2]University of Pennsylvania, Philadelphia, United States, [3]University of Groningen, Groningen, Netherlands, [4]Temple University, Philadelphia, [5]University of California Santa Cruz, Santa Cruz, [6]Center for Coastal Studies, Provincetown, United States, [7]University of Leeds, Leeds, United Kingdom, [8]Arizona State University, Tempe, United States

Abstract: Large body size has evolved at least 11 times in mammals, exemplified by elephants and whales. These species should face a higher lifetime risk of cancer due to the greater probability of oncogenic mutations occurring during somatic evolution in an organism containing 100 to 1000X more cells than a human. However, zoo necropsy data reveals elephants have only a 1-3% probability of death from cancer compared to 11-25% for humans. We find elephant genomes harbor up to 40 alleles of the tumor suppressor gene TP53, and at least some TP53 retrogene copies are transcribed and translated. Functional assays demonstrate TP53 redundancy in elephants is related to an increased apoptotic response to DNA damage in elephant cells when compared to human cells. To investigate cancer suppression in gigantic baleen whales, we have sequenced and assembled the genome of the humpback whale (*Megaptera novaeangliae*). We find that cancer suppression in cetaceans is associated with positive selection on protein-coding genes involved in cell signaling, cell proliferation, apoptosis, as well as cancer. Our findings suggest independent evolution towards gigantism during the mammalian radiation was accompanied by divergent cancer suppression mechanisms, expanding our knowledge of “nature’s toolkit” in fighting tumorigenesis.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-179

Accelerated Gene Turnover and Parallel Gene Family Expansions and Contractions in Marine Mammal Lineages

Claudio Casola [1,*], Tomasz Koralewski [1], Michelle Lawing [1]

[1]Ecosystem Science and Management, Texas A&M University, College Station, United States

Abstract: Colonization of marine environments has independently occurred in three orders of mammals. Comparative genomic studies have shown evidence of both parallel positive selection regimes and nonsynonymous changes in genes of marine mammals and some of these changes were implicated in phenotypic convergence. Gene duplication and gene loss events arguably played a role as well in adaptation to a marine lifestyle for these mammalian taxa, as indicated for example by the loss of multiple keratin genes in cetaceans. Here, we examined parallel expansions and contractions of in gene families in seventeen mammalian genomes, including sequences from four marine mammals: two cetaceans (killer whale and bottlenose dolphin), one carnivore (walrus) and one sirenian (West Indian manatee). Using a maximum-likelihood approach to estimate ancestral states of gene family size implemented in the CAFE program, we found that marine mammal lineages experienced nearly twice as much gene turnover (gene duplications and losses) compared to their sister terrestrial species. Parallel duplications and losses in marine mammal gene families occurred 2.3 and 1.7 times more often than in terrestrial sister species, respectively, after correcting for the total number of expansions and contractions in each lineage. Each pair of marine mammal also showed accelerated gene turnover in comparison to correspondent pairs of terrestrial sister species. A total of 26 gene families exhibited parallel duplications and 63 gene families showed parallel losses in all marine mammals, compared to 7 families with duplications and 47 families with losses in terrestrial sister taxa. Eleven gene families that expanded in marine mammals are involved in regulation of gene expression. Together with expected widespread losses among olfactory receptor and keratin families, gene losses also occurred in processes including iron homeostasis and lipid catabolism. When considering pairs of marine mammal species, we found 617 gene families with convergent gene duplications or losses. These findings underscore the significant role of gene family size evolution in the emergence of convergent phenotypes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-182

Using mutation rate and convergence toward specific mutations to estimate selection and predict evolution of resistance to cancer therapy

Vincent Cannataro [1,*], Carly Stender [1], Stephen Gaffney [1], Ziming Zhao [1], Andrew Greenstein [2], Jeffrey Townsend [1]

[1]Biostatistics, Yale University, New Haven, [2]Gilead Sciences, Foster City, United States

Abstract: Cells within many of our tissues are independently converging towards mechanisms that confer higher cellular fitness—a process that manifests as cancer. Similarly, tumor cells treated with chemotherapy evolve and converge to resistant genotypes. Here, we present a framework that makes use of this convergence to estimate the mutation rate and selection intensity of mutations in somatic tissue using genomic data. We apply our framework to predict the likelihood of—and selection intensity for—different mechanisms of resistance to a therapy in development that targets the *KRAS* G12C mutation. We performed sequencing of 27 *KRAS* G12 positive lung tumors and found that convergence at this site was homogeneous, in that there was no evidence of other oncogenic mutations within *KRAS* or within downstream genes that could confer resistance at the time of treatment. Furthermore, we estimated the *de novo* mutation rate in *KRAS* position 12 and in downstream genes. We found that the mutations likely to confer resistance within *KRAS* and downstream of *KRAS* have similar net mutation rates. Using the mutation rate and the frequency of convergence toward each potential mutation, our approach estimates that *BRAF* V600E mutations would provide the highest fitness advantage for *de novo* resistant subclones. Overall, our findings suggest that resistance to targeted therapy of *KRAS* G12C positive tumors is unlikely to be present at the time of treatment and, among the *de novo* mutations likely to confer resistance, mutations in *BRAF*, a gene with targeted inhibitors presently available, result in subclones with the highest fitness advantage.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-163

Genomics of Rapid Divergence in Lake Malawi cichlids

Cansu cetin [1,*], Patrick danley [1], Aimee howe [1]

[1]Department of Biology, Baylor University, Waco, United States

Abstract: Systems that have experienced replicated convergent evolution provide powerful models to understand genetic basis of complex phenotypes. Cichlid fishes have long been recognized as models providing insight into convergent evolution. However, most examples of convergent evolution in this system are species that have been isolated for millions of years. Here we present evidence of repeated convergent evolution within a recently diverged species rich genus of Lake Malawi rock-dwelling cichlids, the genus *Maylandia*. Two ectomorphs differing in their color pattern, body shape, aggressiveness, and habitat preference have been identified in this genus. A preliminary phylogeny based on RAD-seq data supports the independent evolution of barred and non-barred ectomorphs. Thus these ectomorphs are a valuable system to investigate the genetic basis of a complex phenotype that includes behavioral, morphological, and ecological characteristics. We are interested in the origin and organization of alleles contributing to these phenotypes. In particular, we will test whether convergence has resulted from shared versus novel genetic variation. If it is from shared variation, we will determine if this is from the sorting of ancestral variation or through introgressive hybridization. We will also examine the organization of genes contributing to these phenotypes to assess whether linkage disequilibrium could have facilitated the rapid and repeated divergence of the ectomorphs. Furthermore, QTL influencing many of these characteristics are clustered together on three linkage groups. These findings suggest that the evolution of these ectomorphs may have been facilitated by the genetic origin and organization of the traits involved in their divergence.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-167

Signatures of convergent evolution in chloroplast genes in C4 plants

Tomasz Koralewski [1,*], Claudio Casola [1]

[1]Department of Ecosystem Science and Management, Texas A&M University, College Station, United States

Abstract: Convergence at the phenotypic level has been well documented. Nevertheless, investigations of convergent evolution at the DNA and amino acid levels are fairly recent. In plants, the process of molecular convergence is particularly intriguing in the case of C4 photosynthesis because the C4 pathway has independently evolved multiple times in various lineages of angiosperms. We selected 30 species from grass family, representing both the C3 and C4 photosynthesis types, and investigated evolution of amino acid sequences of 14 orthologous chloroplast genes for signatures of molecular convergence. A given amino acid site was considered convergent if the same amino acid was present in both compared species and a different one in the inferred ancestral sequence, and it was considered divergent if there were different amino acids at the site in both compared species and in the ancestor. To reconstruct ancestral DNA sequences, we used a published phylogeny and the software package PAML. We summarized the cases of convergence and divergence at each amino acid site for each pair of species, and further analyzed the three classes of inter-species comparisons: C3-C3, C3-C4 and C4-C4. Particularly strong signal of convergence in C4 plants came from ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo), a key enzyme in the photosynthesis process. We found signatures of convergence in C4-C4 pairwise comparisons also in two other genes: RNA polymerase beta' subunit and ribosomal protein S18. These preliminary results provide evidence of molecular convergence in multiple chloroplast genes in C4 plants.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-168

A hierarchical Bayesian model for detecting convergent rate changes of noncoding elements on phylogenetic trees

Zhirui Hu [1,*], Jun S. Liu [2], Tim Sackton [2], Scott V. Edwards [2]

[1]Department of Statistics, [2]Harvard University, Cambridge, United States

Abstract: Variation of DNA substitution rates among lineages may indicate gain or loss of function DNA elements. Previous methods estimating branch-specific substitution rates, either by counting the number of substitutions or computing the likelihood ratio between a constant-rate versus variable-rate model on a subtree, assume that at most one change in rate can occur, which is likely unrealistic. Here we introduce a new Bayesian method to model substitution rate changes for conserved non-exonic elements (CNEEs) that can handle multiple losses and gains of conservation during evolution. The model assumes a latent state for each branch on the tree, which can either be neutral, conserved or accelerated, estimates substitution rates for each CNEE, and detects changes of rate by estimating the posterior probability of latent state for each branch. Our method overcomes the challenge of low information content for each CNEE by utilizing hierarchical prior pooling of information from all elements. Furthermore, our method can detect the specific pattern of evolution for a CNEE based on Bayes factor. Simulations show that our method has greater power to detect convergent rate changes in multiple species than likelihood ratio methods. When applied to study convergent accelerations of CNEEs in flightless birds (ratites, Palaeognathae), we identified thousands of CNEEs with increases in substitution rates in one or more ratites but not in other birds, and by identifying branches on which losses of conservation occur demonstrate that these elements putatively lost function in different ratite lineages independently. This method is widely applicable to any set of conserved elements where multiple changes in constraint are expected across the tree.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-165

Utilising Sequence Similarity Networks (SSNs) to Analyse Microbial Evolution by Gene Fusion in Different Species

Yaqing Ou [1,*]

[1]Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, United Kingdom

Abstract: Traditional phylogenetic trees have been using as an effective tool in evolutionary studies for more than a century, although they are only appropriate in describing treelike (divergent) relationships rather than convergent events, such as gene fusions. More recently, sequence similarity networks (SSNs) have been used to analyze divergences and reticulations and to provide a visual and analytical framework for detecting sequence recombination. In a diverse dataset used in this project, a total of 1,111,538 genes were retrieved from public databases. This dataset consisted of the proteomes of 37 eukaryotes, 90 bacteria, 56 archaea and 79 viruses. According to the results of all-versus-all BLAST and using a software program to identify mosaic genes, we found that almost half sequences in this database were composite genes and more than 90% of the species were involved in mosaic events. Furthermore, photosynthetic eukaryotes had significant numbers of fusions, which suggested the association of fused genes function and plant lineages. Also, composite genes are likely to generate new functions with enzymes or biosynthesis. To some degree, our findings supported the irreplaceable place of SSNs in the future evolutionary analysis of gene origin.

Expanded summary*: During the "building" of genes and life in the evolution history, mosaic events are a significant and necessary part. In gene fusion or fission study, two related or unrelated sequences were observed being involved in a convergent event when genes merge to one as well as one gene splits into two genes in a divergent event. However, relying on conventional evolutionary analysis approaches, such as phylogenetic trees, mosaic events will be ignored. In this project, we utilised new technologies named SSNs to investigate gene fusion phenomenon on earth. A large-scale detection and comparison were held between 1,111,538 genes of the dataset comprising of 262 species from three domains of life and viruses. In this database, there are 47.60% genes detected as composite genes, 56.69% genes detecting as component genes as well as 93.51% species taking part in mosaic events. Moreover, gene fusion happened in every species among three domain of life apart from viruses.

Gene fusions have been proved that widespread occurred in all of prokaryotic, eukaryotic, viral and plasmid genome. Moreover, it was estimated that multi-domain genes constituted approximately 40% of the prokaryotic genes and over 67% of eukaryotic genes, which is promising as the results of fusion events. This is consistent with our results that around half genes in the dataset (mainly three domains of life, the virus only occupied 16,775 genes in 1,111,538-sequence dataset) were fused genes.

The extensive existing of mosaic genes is very likely owing to the fact that novel functions are promoted by gene fusion. Although genes remodelling was believed it results in toxic substance or even human cancer, it also played a role in attributing antibiotic-resistance genes in bacteria. From the perspective of evolutionary, fusion genes can not only connect independent and particular functional entities and create a new function from the original one, but also generate new genes with novel functions rather than simply additional functions from the component sequences. What is more, according to the comparison of gene functions between intact dataset and genes in non-transitive triplets, composite genes operated in most aspects, especially in enzyme activity, biosynthesis process and signalling pathway and so forth. The research findings from Wu, Rasmussen and Kellis and Green et al. are helpful to support this result. Wu et al. found there were one-third gene families having rearranged domains in *Drosophila* clade, which considerably influenced the signalling and development process. Also, Green et al. proved that HGT and fused genes facilitated biosynthesis of polyamine enzyme S- adenosylmethionine decarboxylase, aminopropyl transferaseand histidine.

In addition, there are a wide range of reasons contributing to gene fusions: HGT, recombination between species, paralogs sequences communication, intron retrohoming, allopolyploid formation, fractional non-orthologous alternation, various modular operon construction as a result of novel genetic assemblies, novel transposons families, self-governing and incomplete lineage-sorting within alleles and imbalanced ratios of losing character among ancestries.

Among the first 11 species who took up the biggest number of mosaic genes in the database, *Homo sapiens* ranked as the first place and there were four members

pertaining to algae and plants. To some degree, this finding suggested the connection between photosynthetic lineages and foreign composite proteins. The results of Nakamura, Itoh and Martin investigated that the remodellingevents were beneficial for the phylogeny regeneration of plants. Through building proteins similarity networks, Méheust Méheust et al. discovered the wide existence of symbiogenetic genes (S genes) in algal and plant genomes which took effects in improving the capability of oxidative pressure. Those S genes were derived from plasmid endosymbiont and formed the original composite genes by plantae eukaryotes. Even though only the 17 viruses have not been detected taking part in fusedevents, mobile genetic elements, for example, viruses and plasmids, still are considered as crucial parts of non-homologous recombination. Apart from the precious example of S genes from plasmids, Jachiet Jachiet et al. illustrated that in a dataset of three domains of life and mobile genetic elements, there were one-fifth composite genes generating from pieces of viruses and plasmids.

Networks-based methods have been used in many studies in the evolutionary field and the SSNs are helpful when describes the reticular processes. The SSNs will provide an interactive user interface, which can visualise a considerable quantity of protein sequences simultaneously. Nodes and edges were used for representing genes and their biological relationships and hybrid nodes allow the happen of divergent and convergent events, which is hidden in the sole lens phylogenetic trees. From the GDF files and Gephi interface, it is straightforward to observe and analyse every independent gene fusion events. However, the similarity networks will break down when encountering giant connected component (GCC), due to the fact that GCC is too dense to visualisewith Gephi. In conclusion, utilising the new reticular model SSNs can depict complete biological evolutionary relationship, not only the vertical inheritance but also introgression or gene fusions. In total, there were 262 species from three domains of life reconstructed for the dataset in this project, containing more than 1 million genes. After all-versus-all BLAST searching and composite genes identification, there were 529,088 genes defined as fused genes, accounting for 47.60%, in which 371,759 genes belonged to non-transitive triplets. Furthermore, those fused genes distributed among all species of eukaryotes and prokaryotes and most viruses in this dataset. It seems, then, that SSNs, despite having difficult in painting GCCs, will most probably continue to spread and develop as a fresh and advanced technologies of investigating life evolutionary, especially for the extensive existence of gene fusions, in the foreseeable future.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-175

Inference of changes of fitness landscape from sequence data with single-position resolution

Galya V Klink [1,*], Georgii V Bazykin [1 2]

[1]RAS, IITP RAS, [2]Lomonosov MSU, Moscow, Russian Federation

Abstract: Site-specific protein fitness landscape is variable, so that an amino acid at a specific position of a protein confers different fitness values in different species. This variability has been inferred from statistical analysis of pooled data from numerous amino acid positions of large protein alignments. Furthermore, fitness of specific amino acids in different contexts has been measured in directed mutagenesis experiments. However, to our knowledge, no method can detect changes of fitness conferred by an amino acid within a single protein site using sequencing data alone. Here, we develop a method to estimate the probability, for each amino acid at each site, that this amino acid has changed its conferred fitness over the evolutionary time spanned by a phylogenetic tree, using the distribution of substitutions at this site over the phylogeny. Our test performs well on alignments that include hundreds or more species, and allows to compare properties of sites with stable and changing landscapes.

Expanded summary*: Which amino acid confers the highest fitness at a particular position of a protein may change in the course of evolution, and such changes have been implicated to be important in fields ranging from studies of speciation to prediction of the effect of pathogenic alleles in humans. In my PhD project, I analyze sequence data together with phylogenetic trees to study such changes. Using a huge phylogenetic tree of mitochondrial proteins that includes 4350 species of animals and fungi, I have shown that parallel substitutions that repeatedly give rise to the same amino acid at a particular protein position tend to occur in closely related species, while divergent substitutions occur in more distantly related species. This implies that the fitness of different amino acids relative to each other changes in the course of evolution. In a follow-up study, I ask how such changes have affected the lineage leading to our own species. I show that not only the “normal” amino acid variants, but, somewhat unexpectedly, also the amino acids associated with disease in human mitochondrial genes are more likely to arise in parallel in closely related species such as apes, than in more distantly related species such as fruit flies. This implies that even though a variant is deleterious in humans, it is still more benign than other variants that are observed in distantly related species, but probably are lethal for humans.
While such results, as well as results of others, shed light on the mechanisms of evolution, they are based on pooled analysis of large datasets, and tell us little about individual variants. To predict the effects of individual mutations, it is necessary to infer evolutionary changes of fitness conferred by different variants for individual amino acid sites, and the large available amount of sequence data allows us to do this. I have developed an approach to study changes of fitness landscape at a single amino acid position of a protein, and show that the amino acids pathogenic in humans change the fitness conferred by them particularly fast. Such changes may bias methods used to predict effects of mutations by comparative genomics, and should be accounted for.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-173

Dissecting the molecular evolution of toxin insensitivity: insights into the constraints on adaptation.

Lu Yang [1], Santiago Herrera-Alvarez [2], Lucy Cobbs [1], Jamie Ding [1], Maria Del Pilar Rodriguez [2], Ying Zhen [1], Andrew J Crawford [2 3], Peter Andolfatto [1 4,*]

[1]Ecology and Evolutionary Biology, Princeton University, Princeton, United States, [2]Ciencias Biologicas, [3]Museo de Historia Natural ANDES, Universidad de los Andes, Bogota, Colombia, [4]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, United States

Abstract: Despite the central role of adaptation in evolutionary biology, we remain ignorant about numerous aspects of the process, notably: *What developmental, genetic, physiological and population-level factors limit the evolution of new biological traits and the rate of adaptation? How predictable are adaptive evolution outcomes*? Answers to these questions are critical to our understanding of evolution and have important practical implications for how we might expect species to respond to anthropogenic changes. Yet they are difficult to address anecdotally, especially when so few traits have been dissected to the molecular level. Our approach is to examine "convergent" adaptations across distantly related species, and in particular settings in which large assemblages of species have adapted in parallel under a common selective pressure. As a case study, I will discuss our on-going work on the evolution of toxin-insensitivity of Na^+,K^+-ATPase by certain groups of insect herbivores as well as predatory insects and vertebrates. By combining transcriptome sequencing, evolutionary analyses, protein structure modeling and genome engineering, we have revealed the genetic and mechanistic basis for the convergent evolution of cardiac glycoside insensitivity in these groups, and identified factors that constrain its evolution. Beyond the insights into this specific adaptation, our work suggests a powerful new paradigm for studying the adaptive process.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-174

Genetic basis of local adaptation with gene flow in Arabidopsis lyrata

Tuomas Hämälä [1 2,*], Tiina M. Mattila [1], Outi Savolainen [1 2]

[1]Ecology and Genetics, [2]Biocenter Oulu, University of Oulu, Oulu, Finland

Abstract: Differential selection over environments can lead to adaptive divergence when disruptive evolutionary forces have lesser effect. This happens when populations inhabit environments separated by large geographical distances or when selection in contrasting sites is strong enough to overcome the homogenizing effects of the gene flow. The former scenario is a well-studied one, but less is known about local adaptation that occurs despite ongoing gene flow. Here, we aim at quantifying both selection and gene flow in montane populations of the outcrossing and perennial herb, *Arabidopsis lyrata*. Four populations from two altitude gradients at latitudes 61 (from altitudes 300 and 1200 m.a.s.l) and 62 (10 and 1400 m.a.s.l) were reciprocally planted at one of the gradient sites and monitored for two years. Results showed that at both high and low fields, native populations had higher survival and reproductive fitness, consistent with local adaptation. Coalescent simulations using whole-genome resequencing data further indicated that populations from different altitudes have diverged recently and there are significant, but asymmetric, amounts of gene flow between them. Moreover, sequence data are used for divergence outlier scans to search for loci potentially underlying the adaptation in the two separate sites.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-170

Convergent evolution of red blood cell sickling in human and deer

Alexander Esin [1 2], Therese Bergendahl [3], Joseph Marsh [3], Tobias Warnecke [1 2,*]

[1]Institute of Clinical Sciences, Imperial College London, [2]MRC London Institute of Medical Sciences, London, [3]MRC Human Genetics Unit, Edinburgh, United Kingdom

Abstract: Sickle cell anaemia is caused by a single point mutation in the β-globin gene. A small change with large consequences: mutant hemoglobin (HbS) molecules polymerize into tubules upon deoxygenation, coercing red blood cells into the characteristic sickle shape that came to the attention of the medical community in 1910 when Herrick described "peculiar elongated and sickle-shaped blood corpuscles" in a patient with severe anaemia. Remarkably, however, sickled erythrocytes had already been observed 70 years earlier when Gulliver noticed a range of oddly shaped cells, many in the now familiar sickle form, in the blood of white-tailed deer (*Odocoileus virginianus)*. In this talk, I will describe our efforts to determine the genetic basis of sickling in deer, establish whether phenotypic convergence is mirrored at the molecular level and find out whether convergence has been driven by similar ecological pressures (i.e. blood-borne parasites). We determined the sequences of β-globin orthologs from 15 deer species to identify a set of co-evolving, structurally related residues that distinguish sickling from non-sickling deer and suggest a sickling mechanism reminiscent of but distinct from the human disease state. Recurrence of sickling and non-sickling genotypes on independent branches of the deer phylogeny suggest parallel adaptive evolution. I will discuss the role of different evolutionary processes, including introgression, gene conversion, and hemiplasy, in underpinning observed patterns of molecular convergence. Our results illuminate mechanistic, structural, and evolutionary parallels in sickling between primates and cervids, with implications for understanding the ecological regimes and genetic architectures that favour the evolution of sickling.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-171

Dynamic convergent evolution drives the passage adaptation across 48 years' history of H3N2 influenza evolution

Weiwei Zhai [1,*], Hui Chen [1], Qiang Deng [1], Sock Hoon Ng [2], Raphael Lee [3], Sebastian Maurer-Stroh [3]

[1]Human Genetics, Genome Institute of Singapore, [2]DSO National Laboratories, [3]Bioinformatics Institute, Singapore, Singapore

Abstract: Influenza viruses are often propagated in a diverse set of culturing media and additional substitutions known as passage adaptation can cause extra evolution in the target strain, leading to ineffective vaccines. Using 25,482 H3N2 HA1 sequences curated from GISAID and NCBI databases, we found that passage adaptation is a very dynamic process that changes over time and evolves in a seesaw like pattern. After crossing the species boundary from bird to human in 1968, the influenza H3N2 virus evolves to be better adapted to the human environment and passaging them in embryonated eggs (i.e. an avian environment) leads to increasingly stronger positive selection. On the contrary, the passage adaptation to the mammalian cell lines changes from positive selection to negative selection. Using two statistical tests, we identified 19 codon positions around the receptor binding domain strongly contributing to passage adaptation in the embryonated egg. These sites show strong convergent evolution and overlap extensively with positively selected sites identified in humans, suggesting that passage adaptation can confound many of the earlier studies on influenza evolution. Interestingly, passage adaptation in recent years seems to target a few codon positions in antigenic surface epitopes, which makes it difficult to produce antigenically unaltered vaccines using embryonic eggs. Our study outlines another interesting scenario whereby both convergent and adaptive evolution are working in synchrony driving viral adaptation. Future studies from sequence analysis to vaccine production need to take a careful consideration of passage adaptation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

OW-CE2

Causes of parallel molecular adaptation: insights from reverse genetics

Jay Storz*

Abstract: To what extent is adaptive phenotypic convergence caused by parallel changes at the molecular level? Given the typical 'many-to-one' mapping of genotype to phenotype, we don't generally expect replicated changes in phenotype to involve the same underlying mutations. To evaluate the causes of molecular parallelism, we need to figure out which properties distinguish actualized solutions from those of the many nonactualized possibilities. This can be accomplished by using reverse genetics experiments to test the phenotypic effects of individual substitutions (those that actually occurred as well as nonactualized 'might have been' changes). A key finding is that the fitness effects of amino acid mutations are often conditional on genetic background. This context-dependence can reduce the probability of molecular parallelism because it reduces the number of possible mutations that are unconditionally acceptable in divergent genetic backgrounds. Here I evaluate the causes of parallel molecular evolution using experimental data on the molecular basis of hemoglobin adaptation to hypoxia in high-altitude animals. The results reveal that intramolecular epistasis and mutation bias both exert a strong influence on the probability of parallel substitution at the amino acid level.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-180

Evidence of Positive Selection in Parallel Evolution of Closely Related Gammarus Species Genomes

Valentina Burskaia [1,*], Sergey Naumenko [2], Georgii Bazykin [1 2]

[1]Center for Data-Intensive Biomedicine and Biotechnology, Skolkovo Institute of Science and Technology, [2]Institute for Information Transmission Problems, Moscow, Russian Federation

Abstract: We compared the rate of parallel synonymous and nonsynonymous molecular evolution in different systematic groups. Orthologous groups alignments of closely related gammaridae species were used for this issue, and vertebrates exon alignments were added as a reference dataset. Where the same nucleotide substitutions had occurred in more than one lineage, nonsynonymous substitutions were found to be more frequent than synonymous ones. The excess of nonsynonymous parallel substitutions indicates prevalent positive selection at sites of parallel evolution in closely related species, and implies the high role of positive selection in early sympatric divergence.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

OW-CE6

Comparative methods offer powerful insights into social evolution in halictid bees

Sarah Kocher [1,*]

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, United States

Abstract: Natural variation can help us understand how ecological and evolutionary dynamics shape complex traits. Halictid bees or 'sweat' bees harbor extraordinary variation in social behavior. In this group, eusociality has evolved independently 3 times and variation among species encapsulates nearly all forms of social structure, from solitary to primitively eusocial. Importantly, several species exhibit *intraspecific* variation in social behavior on par with that observed between species. In one of these species, *Lasioglossum albipes*, these differences have a genetic underpinning. Population genetic analyses from multiple social and solitary populations have identified a number of sequence polymorphisms associated with social behavior in this species. Further evidence suggests that the behavioral variation observed in *L. albipes* appears to be the result of local adaptation to different environments, similar in many ways to the variation observed between benthic and limnetic populations of sticklebacks. To identify the genetic factors associated with the evolution of social behavior in this group, we have also generated reference genomes for 17 additional halictid species that encompass all of the known gains and losses of eusociality within this family. This enables an integrative examination of the link between the proximate mechanisms underlying variation in social behavior and the ecological processes driving their evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-178

Parallel evolution of toxin resistance in a predator-prey system

Santiago Herrera-Álvarez [1,*], Lu Yang [2], Maria Del Pilar Rodríguez-Ordoñez [1], Andrew J. Crawford [1 3], Peter Andolfatto [2 4]

[1]Department of Biological Sciences, Universidad de los Andes, Bogota, Colombia, [2]Dept. of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, United States, [3]Museo de Historia Natural ANDES , Universidad de los Andes, Bogota, Colombia, [4]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, United States

Abstract: How predictable is genetic evolution? Repeated use of the same genetic changes during *repeated phenotypic evolution* (*i.e.* parallel and convergent evolution) is thought to indicate biases and constraints on the type of mutations that will become fixed in a species lineage. Parallel evolution of toxin resistance in large assemblages of animal is one of the clearest examples of adaptation and represents a useful scenario to test for parallel evolution. Here, we study a natural predator-prey system, that involves the ability of a variety of vertebrate species to feed on toxin-producing toads (Anura: Bufonidae). Bufadienolides belong to the family of endogenous cardioactive steroids that can bind to and inhibit the alpha subunit of the Na^+,K^+-ATPase (ATPα), a protein family affecting muscle contraction, neural function and the membrane potential of cells generally. Using transcriptome data from a number of toad-eating vertebrates, we show that both structural and regulatory changes shape the evolution of toxin resistance and that adaptive evolutionary trajectories are, to some extent, predictable. ATPα is a hotspot for adaptive molecular evolution in species with a cardioactive steroid-rich diet as shown by repeated amino-acid substitutions and parallel changes in gene regulation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

OW-CE4

Detecting signatures of convergent adaptation in genomic data

Kristin Lee [1,*], Jessica Selby [2], John Willis [2], Graham Coop [1]

[1]Department of Evolution and Ecology, University Of California, Davis, Davis, [2]Department of Biology, Duke University, Durham, United States

Abstract: Convergent evolution, when selection for the same trait occurs multiple times independently, can be leveraged as a tool to identify the ecological and molecular basis of adaptation. Thus, it is necessary to be able to accurately identify cases of convergence. We present a composite likelihood-based approach that utilizes genomic data to detect cases of convergent adaptation at the genetic level and the loci involved, taking advantage of the signatures selection has on linked neutral variation. If a population experienced selection recently, we expect the coalescent rates to be elevated around the selected locus, decreasing as function of distance from the selected site. We scan for regions of the genome where these patterns occur in multiple populations showing adaptation in response to similar selective pressures, accounting for neutral population structure. Additionally, we have developed coalescent theory to model convergence due to multiple origins of the beneficial allele or a single origin from either gene flow or selection on shared ancestral standing variation. We can distinguish among these modes as well as cases that are not truly convergent, such as selective events that were shared in the ancestor of presently isolated populations or migration and drift simply increasing the frequency of an allele selected for in another population. We illustrate our method with a novel application to genome-wide polymorphism data from three populations of *Arabidopsis lyrata* that show convergent adaptation to serpentine soils.

Expanded summary*: Convergent adaptive evolution demonstrates the impressive ability of natural selection to repeatedly shape phenotypic diversity and gives insight into how evolution is constrained by molecular pathways. It can also be utilized as a tool to help identify the ecological and molecular basis of adaptation. Thus, it is necessary to be able to accurately identify cases of convergence from those due to chance events. I focus here on convergence that occurs at the genetic level, with geographically separated populations converging to the same phenotype through parallel genetic changes at the same loci. Genomic data and knowledge of linked variation allows us to address novel questions related to convergence. We take advantage of the signatures of selection at linked neutral sites to detect cases of convergent adaptation and the loci involved.

I make use of the fact that hitchhiking can be modeled as an increase in the variance in neutral allele frequencies around a selected site within a population. Selected alleles present in multiple populations at the same locus can have multiple independent mutational origins or a single origin. In the latter case, adaptation in different populations may proceed by means of selection on the same ancestral standing variation, or a single allele spread throughout the populations via gene flow. In these single origin cases, hitchhiking events that are shared between populations act to increase covariance between populations in neutral allele frequencies at loci near the selected site. The patterns of this covariance, and its decay with the distance away from the selected site, are determined by the mode of convergence. I have developed coalescent theory to specify such covariances for various models of migration and shared ancestral variation. Distinguishing between these modes will elucidate how often adaptation is limited by mutational input and the role of standing variation and gene flow in adaptation. We can also model cases where populations appear to be but are not truly

convergent. For example, selective events that were shared in the ancestor of presently isolated populations or migration and drift simply increasing the frequency of an allele selected for in another population.

I incorporate these hitchhiking effects, both within and between populations, into a multivariate normal model of allele frequencies under population structure, accounting for covariance due to shared drift. Based on this model, I develop a composite likelihood-based method that utilizes genomic data to identify cases of true convergence and loci involved, as well as distinguishing among different modes of convergent evolution across multiple populations. This flexible framework allows us to identify modes of convergent adaptation for many populations with arbitrary relationships. It also allows us to distinguish between more subtle models, enabling the identification of the origin and the direction of gene flow of a beneficial allele.

This method should be of wide use as many investigators are sequencing population genomic samples from across the geographic range of species. I have previously applied this method to regions that we *a priori* presumed to be under convergent selection to distinguish between the modes of convergence in the adaptation in various datasets. Here, we extend the method to scan for regions of the genome under convergent selection, with a novel application to genome-wide polymorphism data from three populations of *Arabidopsis lyrata* exhibiting tolerance for serpentine soils in the Northeastern United States.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-176

Is specialization driven by loss-of-function? A comparison of chemoreceptor evolution in a generalist/specialist pair of pine sawflies.

Kim Vertacnik [1,*], Catherine Linnen [1]

[1]Biology, University of Kentucky, Lexington, United States

Abstract: One out of every four species is a plant-feeding insect, and most are specialized for particular plant taxa. Because they determine an insect's ability to detect host plant compounds, chemosensory proteins may play an essential and repeatable role in host preference evolution. Additionally, a large body of work in *Drosophila* suggests that diet breadth reductions may be predictably driven by loss-of-function (LOF) mutations in chemosensory genes. However, because most *Drosophila* specialists are also island endemics, it is difficult to disentangle the contributions of genetic drift and natural selection to chemosensory gene evolution and LOF mutation accumulation. Here, we examine the role of chemoreceptor LOF mutations in a pair of pine sawfly species that are widely distributed across eastern North America, but differ in diet breadth. While *Neodiprion pinetum* is a specialist on white pine (*Pinus strobus*), *N. lecontei* is a generalist that feeds on many different pine species, but avoids white pine. We first present data demonstrating that the specialist has increased preference, improved oviposition performance, and improved larval feeding performance on white pine. To test the hypothesis that LOF mutations contribute to white pine specialization, we assembled both genomes de novo and manually annotated the olfactory and gustatory receptor gene families in each genome. Using these data, we evaluated whether rates of pseudogenization and non-synonymous substitutions were higher in the specialist. Our results indicate that although host specialization is a complex trait, host-use evolution may be predictable at the level of gene family and gene function.

Expanded summary*: Over one quarter of all described species are plant-feeding insects. This impressive biodiversity is driven by extensive specialization and adaptation to diverse host taxa. The overall goal of my research is to determine the extent to which the genetic mechanisms underlying host adaptation are predictable. Specifically, I focus on genome-wide changes associated with two types of convergence in host-use phenotypes: (1) convergent adaptation to the same host taxa, and (2) convergent shifts between generalist (many hosts) and specialist (few hosts) lifestyles.

To increase our understanding of the genetic basis of biodiversity and evolutionary repeatability, I am working on a comparative study of host preference gene families in a clade of 19 pine sawfly species that vary in the number and identity of their preferred pine species. Preference is a behavioral phenotype describing the selection or ranking of specific plant taxa. These decisions rely on taste and smell, which are mediated in part by chemoreceptor proteins (gustatory and olfactory receptors). Previously, comparative work on chemoreceptor gene evolution in *Drosophila* demonstrated that shifts from generalist to specialist lifestyles are accompanied by loss-of-function (LOF) mutations and elevated rates of non-synonymous substitutions in chemosensory gene families. Although these findings suggest that the genetic basis of specialization may be predictable, they are confounded by demography because the best studied *Drosophila* specialists are often island endemics.

To determine if the genetic correlates of specialization in *Drosophila* also characterize other herbivorous insects, I am exploring chemosensory gene evolution in an eastern North American clade of *Neodiprion* sawflies (Hymenoptera). This is a monophyletic group of 19 pine-feeding species with repeated shifts between generalist/specialist lifestyles and repeated gains/losses of the same pine hosts. Additionally, these species have a wealth of life history data, and broad, overlapping geographical ranges such that specialization is not confounded by demography. In this system, I am testing the hypothesis that host-use changes involve predictable genetic mechanisms, specifically that: (1) reductions in diet breadth are accompanied by LOF mutations and elevated d_N/d_S in chemosensory gene families, and (2) independent host gains/losses of a particular host are associated with changes at the same loci. To test these predictions, I am performing two comparisons. The first is between sister taxa that differ in both preferred host and diet breadth. Whereas *N. pinetum* is a white pine specialist (*Pinus strobus*), *N. lecontei* is a generalist on multiple pine species, but avoids white pine. I have assembled a draft genome for each species and manually annotated chemoreceptor genes for an in-depth comparison of gene loss, pseudogenization, and signatures of positive selection (d_N/d_S). The second comparison is across the entire eastern North American *Neodiprion* clade. For each species, I am assembling a hybrid genome (reference-based + de novo) and manually annotating the chemoreceptor genes. Due to repeated shifts to the same hosts and independent changes in diet breadth, I can

evaluate the repeatability of patterns identified in *Drosophila* specialists, and assess genetic convergence associated with the gain and loss of the same pine hosts.
For this SMBE meeting, I will have data from the in-depth generalist/specialist comparison. I am excited to share this work and get ideas for a future project identifying signatures of selection accompanying divergence in host use among populations *within* species. What I can contribute is my expertise in developing genomic resources for non-model systems, comparative genomics, and the genetic basis of adaptation. The broad significance of my research is that it addresses fundamental questions about the predictability of evolution—specifically mechanisms underlying ecological specialization—which in turn will improve our understanding of species formation and biodiversity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

POB-177

Population genomics of herbicide resistance and agricultural adaptation in Capsella bursa-pastoris

Julia Kreiner [1,*], Stephen Wright [1], John Stinchcombe [1]

[1]Ecology & Evolutionary Biology, University of Toronto, Toronto, Canada

Abstract: Adaptation of ruderal plant populations occurs rapidly in response to agricultural regimes, despite many being introduced species with complex demographic histories. In particular, weed populations exposed to herbicide use experience extreme and predictable selection, recurrent population bottlenecks and expansions, and for *Capsella bursa-pastoris*, a history of widespread colonization. Acetolactate synthase resistance has been reported for *C. bursa-pastoris*, yet we know little about the population genomics of resistance adaptation and demography. To investigate demographic history and population structure, we genotype-by-sequenced 96 individuals from the Canadian Prairies and compared diversity to 261 Eurasian individuals. We found that Canadian agricultural populations originate from Northern Europe, likely through multiple colonization events. Demographic models estimated a strong genetic bottleneck of *C. bursa-pastoris* during its introduction to Canada. We gauged the frequency and number of resistance mutations present across the Canadian Prairies by pool-sequencing whole genomes from 192 individuals. Herbicide resistance is unevenly distributed across the range and has independently evolved at least three different ways, one of which was previously only characterized in experimental populations. Both subgenomes of the tetraploid, *C. bursa-pastoris*, harbored mutations for resistance. By comparing genetic diversity and the population branch statistic between agricultural Canadian and natural European populations, we identified candidate genes putatively involved in agricultural adaptation with functions including detoxification, stress response, growth, and flowering time. Fundamentally, this work provides a unique opportunity to identify how the genetic basis of adaptation is shaped by demography, the nature of these evolutionary processes, and how weeds persist under strict agricultural regimes.

Expanded summary*: Adaptation, the process by which organisms become better suited to their environment, is fundamental for species formation and persistence[1]. Despite this, the rate and population genetics of adaptation remains poorly understood. A growing number of examples of contemporary evolution provide new opportunities to investigate the ongoing dynamics of adaptation in real time. Notably, adaptation of natural plant populations occurs rapidly in response to agricultural regimes, despite many being introduced species with complex demographic histories. Crop cultivation provides weedy species with the opportunity to colonize and adapt to agronomic conditions, negatively impacting crop harvests and leading to a novel evolutionary arms race with crop protection measures[2]. This phenomenon has received little attention from molecular and evolutionary biologists, yet is of dire global importance for food security and species conservation in our changing world.

The evolution of resistant agricultural weed species in response to single target herbicides represents a remarkable case of convergent phenotypic evolution, with over 450 unique cases documented[3]. Resistance to acetolactate synthase (ALS) inhibitors was first reported in 1968 and since then, has been well characterized with many point mutations conferring resistance identified across the ALS gene[4]. Yet, it remains that we know little about the frequency and number of independent resistance mutations occurring across agriculturally important regions. By studying the occurrence of beneficial mutations in the context of herbicide resistance, we can also learn about the relative roles of soft and hard selective sweeps during adaptation. Although extreme selective pressures are prevalent in weed-herbicide systems, demographic history also plays an important role in shaping the genetic diversity of herbicide resistant weeds. Resistant weeds have recently gone through recurrent population bottlenecks and expansions between herbicide applications, but many of these species are introduced and therefore have also experienced these events on a broad continental scale.

As one of the most globally widespread weeds, *Capsella bursa-pastoris* has repeatedly been reported as herbicide resistant. *C. bursa-pastoris* is a selfing, allotetraploid formed approximately 150,000 years ago from the diploids *C. grandiflora* and *C. orientalis* where it was first introduced to North America in the 17th century[5,6]. To investigate demographic history and population structure, we compared genotype-by-sequence data from 96 individuals across the Canadian Prairies to Eurasian samples. We find that Canadian agricultural populations originate from Northern Europe, likely through multiple colonization events, confirming previous work[6] on a whole-genome scale. Coalescent models of Canadian *C. bursa-pastoris* since its split from European populations gives evidence that the species experienced a strong genetic bottleneck during its introduction to Canada. We gauged the frequency and number of

resistance mutations present across the Canadian Prairies by sequencing genomes from 192 individuals in putative low and high resistance pools. We find that herbicide resistance is unevenly distributed across the range, with an average of 6.5% in the east (low) and 17% in the west (high). Resistance has independently evolved at least three different ways, one of which was previously only characterized in experimental populations[7]. Resistance mutations were present on both subgenomes, suggesting that for partially dominant mutations such as ALS resistance, polyploidy may speed the rate of evolution by increasing the mutational target size. By comparing genetic diversity and the population branch statistic between agricultural Canadian and natural European populations, we looked for signals of hard and soft selective sweeps and identified candidate genes putatively involved in agricultural adaptation with functions including detoxification, stress response, growth, and flowering time.

Fundamentally, this work provides a unique opportunity to identify how the genetic basis of adaptation is shaped by demography, the nature of these evolutionary processes, and how weeds persist under strict agricultural regimes. Ultimately, these results will inform farmers of costs and benefits of different weed control measures, and improve the adoption of integrated weed management.

(1) Bell 2013. *Philos Trans R Soc Lond B Biol Sci*, *368*(1610); **(2)** Neve et al., 2009. *NewPhyt*, *184*(4); **(3)** Heap 2017. www.weedscience.com; **(4)** Tranel, & Wright 2002. *Weed Sci*, *50*(6); **(5)** Douglas et al., 2014. *PNAS*, *112*(9); **(6)** Neuffer & Hurka 1999. *Mol Ecol*, *8*(10); **(7)** Ott et al., 1996. *J Mol Biol*, *263*(2).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Convergent evolution

OW-CE1

Convergence and reversion in the presence of gene tree discordance

Matthew Hahn*

Abstract: Phenotypic convergence is often an outcome of adaptive evolution, when different species find similar solutions to the same problem. Unraveling the molecular basis of convergence provides a way to link genotype to adaptive phenotypes, but can also shed light on the extent to which molecular evolution is repeatable and predictable. Recent genome-wide studies have looked at both convergence and reversion—when a second substitution takes a lineage back to the ancestral state—across a wide range of taxa. A striking pattern uncovered in these studies is that both convergence and reversion diminish over time, with less closely related species having proportionally fewer such substitutions. This pattern has been ascribed to intramolecular epistasis constraining the space that proteins can explore. In this talk I consider gene tree discordance as an alternative cause of changes in convergence and reversion over time. I explain how gene tree discordance—either caused by biological or technical factors—can produce similar patterns on its own, and that the patterns in many published datasets seem to be explained by discordance. I present multiple ways to distinguish true convergence and reversion from the changes mimicked by discordance, including a likelihood ratio test for individual substitutions of interest. These results highlight the important (but insidious) role gene tree discordance can play in conclusions about molecular evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE2

Epigenetic Regulatory Loci of Human Brain Evolution and their role on the susceptibility to neuropsychiatric disorders

Isabel Mendizabal [1 2,*], Lei Shi [3 4], Thomas E. Keller [1], Genevieve Konopka [5], Todd M. Preuss [6], Tzung-Fu Hsieh [7], Enzhi Hu [8 9], Zhe Zhang [10], Bing Su [8], Soojin V. Yi [1]

[1]School of Biological Sciences, Institute of Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, United States, [2]Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country, Leioa, Spain, [3]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China, [4]The Molecular & Behavioral Neuroscience Institute,, University of Michigan, Ann Arbor, [5]Department of Neuroscience, University of Texas Southwestern Medical Center, Dallas, [6]Yerkes National Primate Research Center, Emory University, Atlanta, [7]Department of Plant and Microbial Biology and Plants for Human Health Institute, North Carolina State University, Raleigh, United States, [8]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, [9]Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing, [10]Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

Abstract: Human brains underwent an unparalleled degree of recent evolutionary innovation at several levels. We are interested in understanding how epigenetic divergence, in particular DNA methylation difference, might have contributed to human brain evolution. We use a comparative methylomic approach in which whole genome DNA methylation maps of human brains are compared to those of chimpanzees and rhesus macaques. The rhesus macaque brains provide outgroup information to identify potentially human brain-specific changes of DNA methylation. Using this method and using fairly conservative criteria, we identified hundreds of genomic regions that show human brain specific DNA methylation patterns, which we refer to as 'differentially methylated regions (DMRs)'. Among these DMRs, we tested 38 human-specific DMRs using targeted deep genomic and bisulfite sequencing in an independent panel of 37 individuals from five primate species, and found that our comparative methylomic approach reliably identifies potentially human-specific DMRs. These DMRs show not only DNA methylation changes but also signatures of histone H3K4me3 modification. We also identified several 'clusters' of human specific DMRs that are known to be involved in active chromatin loops. Together, these findings indicate that substantial epigenetic reprogramming has occurred during human brain evolution, with implications for regulation of transcription. We are further exploring the impact of human brain specific epigenetic changes to the prevalence of some neuropsychiatric diseases.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE4

Genetic conflict shapes Drosophila telomere biology

Bastien Saint-Leandre [1], Yuh Chwen G. Lee [2], Mia Levine [3,*]

[1]Department of Biology, University of Pennsylvania, Philadelphia, [2]Genome Biology, Lawrence Berkeley National Lab, Berkeley, [3]Department of Biology and the Epigenetics Institute, University of Pennsylvania, Philadelphia, United States

Abstract: Virtually all eukaryotes rely on telomerase to maintain chromosome length. Fruit flies are a widely studied exception. Instead of telomerase, Drosophila relies on domesticated retrotransposons that insert almost exclusively in telomeric DNA. Although hailed as an exemplary "genomic symbiosis," sporadic accounts of rapid evolution in both the telomere packaging proteins and the transposable element arrays to which they localize implicate intra-genomic conflict. Consistent with a molecular arms race model, we have documented pervasive positive selection and recurrent gene turnover at Drosophila telomere protection genes. Those genes restricted to only a few lineages are frequently expressed primarily in the germline—precisely where we expect intra-genomic conflict to play out. Moreover, we have documented exceptionally rapid turnover of telomere-specialized transposable elements across only a few million years of Drosophila evolution. This recent and exceptionally rapid evolution of both telomeric retrotransposons and end-protection proteins suggest that, across Drosophila evolution, telomere integrity requires a recurrently re-negotiated truce between host proteins and telomere lengthening elements. To test this model, we leverage CRISPR/Cas9-mediated transgenesis to cleanly swap into *D. melanogaster* highly diverged alleles encoded by its close relatives. For one fast evolving telomere end protection gene, the non-*melanogaster* allele fully complements the mutant phenotype—we observe no catastrophic end-to-end chromosome fusions. The same genotype, however, hypertranscribes *D. melanogaster's* telomere-specialized retrotransposons. This failure to complement transposon silencing is consistent with an ongoing molecular arms race shaping Drosophila telomere biology and reveals the genetic determinants of molecular domestication (and re-domestication) over evolutonary time.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-41

Diversity and evolution of transposable elements in the Arabidopsis lyrata and A. halleri genomes

Sylvain Legrand [1,*], Thibault Caron [1], Florian Maumus [2], Eléonore Durand [1], Sophie Gallina [1], Lucie Huyghe [1], Sol Schvartzman [3], Marc Hanikenne [3], Vincent Castric [1]

[1]Evo-Eco-Paleo, University of Lille, Lille, [2]URGI, INRA, Versailles, France, [3]Department of Life Sciences, University of Liège, Liège, Belgium

Abstract: Transposable elements constitute the major part of most eukaryotic genomes. The proliferation of these genetic selfish elements is deleterious in many cases but also highly contributes to genome evolution. Sequencing of the outcrossing plant *Arabidopsis lyrata* [1] revealed an important population of recently inserted transposable elements that is not observed in the closely related *A. thaliana* genome, indicating either more successful recent transpositions in *A. lyrata* or more successful elimination of recent TE insertions in the predominant selfer *A. thaliana*. The mechanisms that led to this burst remain unclear and may include either lower efficacy of silencing of TEs and/or lower deleterious effects of TE insertions near genes in *A. lyrata* than in *A. thaliana*. To answer these questions, we characterized the dynamics of TEs at an even finer micro-evolutionary scale by comparing the TE contents in two genome assemblies of the sister species of A. lyrata, A. halleri, that diverged less than one million years ago. Using a hybrid approach combining long (PacBio) and short (Illumina) reads, we obtained an *A. halleri* genome assembly composed of 3,152 scaffolds. Annotation of TEs was performed in parallel in both genomes to avoid ascertainment bias [2], and the distribution of TE families showed that Gypsy, Line, Copia, helitron, and MuDR are predominant in both species. Based on patterns of sequence divergence within families, we clearly confirmed that *A. halleri* also contains the two separate populations of TEs with either 'recent' or 'old' insertion histories that had been identified in *A. lyrata*. Interestingly, however, we found very few TEs at orthologous positions, suggesting very rapid turnover. Hence, although the two outcrossing species seem to share a common evolutionary dynamic of TEs, which is very different from that in the selfer *A. thaliana,* the comparison indicates a very rapid process of insertion and deletion, such that their two populations of TEs have already become largely distinct. We are currently examining the impact of gene density, chromosomal location, TE length, small RNA production and methylation patterns on the dynamics of TE insertions and deletion. We are also using population frequencies of TE insertions to evaluate the strength of natural selection.

1. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, et al. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat. Genet. 2011;43:476–81.
2. Flutre T, Duprat E, Feuillet C, Quesneville H. Considering Transposable Element Diversification in De Novo Annotation Approaches. PLOS ONE. 2011;6:e16526.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE11

Behavioral Variation in Primates as a Function of Epigenetic Regulation and Genetic Polymorphism at the Oxytocin Receptor (OXTR) Gene

Christina Barr [1,*], Carlos Driscoll [1], Stephen Lindell [1], Isaac Miller-Crews [1], James Kehler [2], Raja Kittappa [1], Dee Higley [3], Steve Suomi [4]

[1]Lab of Comparative Behavioral Genomics, NIH/NIAAA, Rockville, [2]NIDDK, NIH, Bethesda, [3]Department of Psychology, BYU, Provo, [4]Laboratory of Comparative Ethology, NIH/NICHD, Bethesda, United States

Abstract: Adaptations to stressful environmental conditions can occur at both species and individual levels. While the former occurs through genetic polymorphism, the latter can occur through epigenetic processes and may be a conduit for informing offspring of environmental challenge. In rodents, disruption in maternal care (DMC) influences epigenetic regulation in stress-vulnerable areas of brain. Here, we employed ChIP-SEQ for H3K4me3 (a modified histone which marks "active" promoters) to examine effects of early environment using archived rhesus macaque (*Macaca mulatta*) hippocampus from animals that had experienced DMC (Higley). There was significant variation of H3K4me3 binding at genes critical to behavioral stress response, the most robust being the oxytocin receptor gene (*OXTR*), at which we observed DMC-related decreases in H3K4me3 binding and a corresponding decrease in RNA expression. We wanted to determine whether there were gain-of-function *OXTR* polymorphisms that could partially "rescue" the DMC behavioral phenotype (archived dataset, Higley). Sequencing the regions encoding amino acid residues known to be critically involved in binding of oxytocin to the receptor, we identified a non-synonymous SNP. This SNP predicted increased responsivity to exogenously administered intranasal oxytocin, suggesting a gain-of-function role (Suomi). Prior studies show that DMC infants exhibit higher levels of reactivity to social separation. We demonstrated that DMC infants carrying this *OXTR* allele do not exhibit as great of a separation response. These data indicate that the oxytocin system is involved in social separation response and suggest that epigenetic down-modulation of *OXTR* could contribute to behavioral differences observed in DMC animals. To see if such findings could translate across species, we queried the human *OXTR* gene (UCSC) and found indication of both between species conservation and regulation of *OXTR* by epigenetic processes (both DNA- and Histone-related). These observations suggest that purifying selection is acting at this region and that there is a potential role for epigenetic processes in determining individual responses in humans. Epigenetic changes at *OXTR* may represent predictive adaptive responses (aka, increased sensitivity to environmental challenge or an insecure attachment style) that could impart readiness to respond to environmental challenge or maintain proximity to a caregiver, but also contribute to behavioral pathology. Our data in macaques demonstrate that there is a gain-of-function *OXTR* polymorphism that permits animals to partially overcome the behavioral effects early maternal deprivation, which could have translational implications for human disease and personality disorders. It is possible that epigenetic regulation and *OXTR* polymorphism act alone or interactively to predict species-level differences in social interactions and stress response, and we are currently working toward developing *in vitro* systems to determine molecular mechanisms controlling oxytocin receptor function as they persist or differ across species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-39

Genomic landscape of cytosine methylation dependent mutations in Vibrio cholerae

Supriya Khedkar [1,*], Mohak Sharda [1], Aswin Seshasayee [1]

[1]National Centre for Biological Sciences (NCBS), Bangalore, India

Abstract: Epigenetic modifications play a key role in gene regulation and in recognition of self DNA in bacteria. In-spite of their positive role in cell survival, modifications like cytosine methylation incur a mutational cost. Cytosine methylation, specifically 5-methylcytosine (5mC), is prone to hydrolytic deamination which leads to C → T and G→ A transitions. Here, we first study the abundance of mutagenic cytosine methylation target motifs and show that bacteria like *Vibrio cholerae* might use motif avoidance as a strategy to minimize the mutational effect of deamination of methylated cytosine. Second, by performing SNP analysis on whole genome sequence data from *Vibrio cholerae* patient isolates we show a) high abundance of cytosine methylation-dependent mutations in the cytosine methylation target motif RCCGGY, b) 95% of these C → T and G → A transitions in the coding region lead to non-synonymous substitutions and c) many of these transitions are associated with membrane proteins and are implicated in virulence. Thus, our SNP analysis of *Vibrio* genomes implicates the role of cytosine methylation in generating genotypic diversity with an adaptive potential.

Expanded summary*: Epigenetic variation is widely studied in the context of its gene regulatory capacities mostly in eukaryotes and recently in bacteria (Walsh & Xu, 2006). While the field of eukaryotic epigenomics is dominated by 5-methylcytosine (5mC) at CpG islands (Deaton & Bird, 2011), bacteria show epigenetic variations like 6mA, 4mC and 5mC (Blow et al., 2016). Traditionally epigenetic variation in bacteria was known only in the context of bacterial Restriction Modification (RM) systems, but increasing evidence suggests the role of adenine and cytosine methylation in regulating gene expression. Specifically, cytosine methylation has been recently implicated in modulating envelope stress response in *Vibrio cholerae* (Chao et al., 2015). However, in-spite of the positive role of cytosine methylation in regulating gene expression, there is a high mutational cost associated with methylated cytosine. This is due to: a) high frequency of spontaneous deamination of methylated cytosines that lead to C → T and G → A transitions (Shen, Rideout & Jones, 1994) and b) absence of dedicated mechanisms to repair these transitions (in *Vibrio cholerae*) during the stationary phase (the phase that comprises majority of the bacterial life cycle in the wild). Thus, methylated cytosines are mutational hotspots that introduce C → T and G → A transitions within the bacterial genome, but the effect of these mutations on bacterial phenotype and genome composition is unknown. To address this, we determined the abundance and predicted the consequences of cytosine methylation associated mutations within *Vibrio cholerae* genomes.

In *Vibrio cholerae*, cytosine methyltransferase (VchM) methylates the first cytosine within its target motif RCCGGY (R – A/G and Y – C/T). We found that the VchM target motif (RCCGGY) was underrepresented in *V. cholerae* genome as compared to random expectation. But in-spite of this under representation we observed ~2000 instances of the target motif across the *V. cholerae* genome. Moreover, our previous analysis had predicted that 97% of these 5mC → T transitions within the target motifs would lead to non-synonymous substitutions. We found that this prediction was reflected in the whole genome sequence data of *Vibrio cholerae* patient isolates (isolated during different cholera pandemics). Our SNP (Single Nucleotide Polymorphism) analysis revealed that 95% of the identified cytosine methylation based transitions within *V. cholerae* patient isolates were non-synonymous and 20% of these were associated with virulence associated genes. We also found that the target motif was enriched in membrane associated genes and bacterial response regulators, implicating the potential role of these SNPs in adaptation to the host environment and in niche expansion.

Thus, the genomic landscape of 5mC → T transitions described by our computational analysis identified consequential mutations within some of the virulence associated genes. Experimental studies on these mutants is called for. Specially, since *V. cholerae is* a major disease causing pathogen and one of the leading cause of death worldwide. Importantly, our analysis showed how the presence

of a 5-methylcytosine methyltransferase affects bacterial genome composition by introducing C → T and G → A transitions and through selection acting towards avoidance of the target motif within the genome. Finally, we think that the bacterial species harboring multiple methyltransferases (*Helicobacter* and *Neisseria*), identified in our analysis (not shown) could act as excellent model systems to study the effect of epigenetic modifications on genome composition.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE13

Long Interspersed Nuclear Elements and Monoallelic expression

Clara Pereira [1,*], Aoife McLysaght [1]

[1]Smurfit Institute of genetics, Trinity College Dublin, University of Dublin, Dublin, Ireland

Abstract: Monoallelic expression (MAE) is an intriguing epigenetic process because it renounces to the advantage of diploidy, thought to have evolved to buffer the effects of mutation. X-chromosome inactivation (XCI) is a notable example of MAE. It provides X-linked genes dosage balance with respect to the autosomes in female mammals. The high density of Long Interspersed Nuclear Elements (LINE) across the X-chromosome has an important role in the XCI mechanism: old inactive LINE sequences, highly enriched in the X chromosome, help the spread of silencing molecules, and heterochromatinization, while younger, intact and active LINE elements contribute to the mobilisation of otherwise prone to escape X-linked genes to the heterochromatic region. Here, we explore the hypothesis that the autosomal MAE genes may also be affected by LINE elements or other repetitive elements to promote mono-allelic silencing. This hypothesis has been proposed before, but despite the subsequent detection of genome-wide monoallelic expression, it has not been further reported. We have used this new wealth of data and bioinformatic tools to search for genomic evidence of enrichment of LINE and other repeat elements in the vicinity of MAE genes. Our results show a consistent but modest increase in LINE density in the flanking region of autosomal MAE genes. The distribution of LINE elements also shows that some families of monoallelic genes may rely on the presence of LINE elements but not others, suggesting different regulatory mechanisms contributing to MAE.

Expanded summary*:

Random autosomal monoallelic expression (MAE) is an epigenetic phenomenon that controls the relative expression of maternal and paternal alleles in thousands of mammalian genes1. Despite MAE's widespread character2, the molecular mechanisms involved in its establishment and maintenance remain unknown1. Recent work shows that the sets of genes subject to MAE are highly consistent across individuals3, and conserved between the human and mouse species4. The extended aim of my work is to test wether such conservation is due to specific regulatory elements in the genome associated with MAE genes. The suggestion that genetic determinants underlie random monoallelic expression was first proposed upon observations of X-autosome translocation studies, and enrichment of Long Interspersed Nuclear elements (LINE) in the non-escapee regions of the X-chromosome12,13. Parallels of XCI and MAE have been long evident, and a significant correlation between MAE and LINE enrichment was drawn14,15. A role for these relics of repetitive transposable elements in establishing or maintenance of autosomal MAE is yet to be defined, but it's supported by the enrichment of repetitive sequences around genes showing asynchronous replication16, an epigenetic state correlated with monoallelic expression. It seems thus that a first approach to the detection of a genetic component of MAE is to solve this lingering question of whether LINE elements are enriched in the vicinity of autosomal MAE genes. Our ongoing analysis on the mouse, human and chimp genomes may help identify distinct groups of MAE genes that are controlled by different silencing processes, setting the stage to further discovery of the models of genetic control of MAE. Understanding monoallelic expression may unlock important knowledge on the dynamics of the genome, its evolution, and understand variation in disease outcomes.

1. Gendrel et al., "Random Monoallelic Expression of Genes on Autosomes: Parallels with X-Chromosome Inactivation.," Sem Cell & Dev Biol 56 (2016).
2. Gimelbrant et al., "Widespread Monoallelic Expression on Human Autosomes.," Science 318, 5853 (2007)
3. Savova et al., "Genes with Monoallelic Expression Contribute Disproportionately to Genetic Diversity in Humans.," Nat Genet 48 (2016).
4. Nag et al., "Chromatin Signature Identifies Monoallelic Gene Expression Across Mammalian Cell Types.," G3 (2015).
5. Lyon, "X-Chromosome Inactivation: a Repeat Hypothesis.," Cyt Cell Genet 80, 1 (1998).
6. Bailey et al., "Molecular Evidence for a Relationship Between LINE-1 Elements and XCI: the Lyon Repeat Hypothesis.," PNAS 97, 12 (2000): 6634–39.
7. Allen et al., "High Concentrations of LINE Sequence Distinguish Monoallelically Expressed Genes," PNAS 100,17 (2003): 9940–45.
8. Gendrel et al., "Developmental Dynamics and Disease Potential of Random Monoallelic Gene Expression," Dev Cell 28, 4 (2014).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-40

Differences in transposable element expression between sexes in Drosophila

Lauren Gibilisco [1,*], Doris Bachtrog [1]

[1]Integrative Biology, UC Berkeley, Berkeley, United States

Abstract: Transposable elements ("TEs") are mobile genetic elements whose movement can be deleterious to their hosts. Host genomes have evolved multiple strategies to suppress TE mobilization, including targeting by small RNAs and heterochromatin formation. In the germline and during early embryogenesis, TE repression is particularly important because any activity can cause new, heritable insertions. We used genomic reads to annotate repeats, most of which map to known TEs, in two fly species, *Drosophila pseudoobscura* and *D. miranda*. We used RNA-seq data from eight embryonic stages (stage 2 - stage 12) to characterize repeat expression during the maternal-to-zygotic transition in male and female embryos and found higher repeat expression during early development in males compared to females in both species. We hypothesize that the Y chromosome influences heterochromatin establishment by acting as a “sink” for repressive chromatin marks, causing a delay in heterochromatin establishment at repetitive sequences in males compared to females. We are currently performing ChIP-seq on single embryos to quantify differences in heterochromatic marks at repeats in males compared to females. We are also sequencing genomic DNA from single embryos and parental gonads from single female x single male crosses to quantify de novo insertions in males.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-43

Comparative Analysis of DNA Methylation from Whole Genome Bisulfite Sequencing Maps of Hymenopteran Insects

Brandon Smith [1,*], Xin Wu [1], Soojin Yi [1]

[1]Georgia Institute of Technology, Atlanta, United States

Abstract: Methylation of nucleotides in a CpG context has been implicated in multiple regulatory processes in many biological systems. In particular, discovery of functional, DNA methylation systems in many insect genomes provides an opportunity to understand the fundamental role of DNA methylation in deep phylogenetic scale. One of the most significant differences between the genomic methylation patterns of insects and the well-studied mammals is that in the former, DNA methylation is generally limited to a small portion of the CpG's in the whole genome. Also, in mammals, methylation of CpG islands has been correlated with repressed expression while, in insects, levels of DNA methylation in CpG's often correlates with increased expression. This fundamental difference suggests that tools that have been developed based upon mammalian patterns cannot be directly applied to the analysis of insect DNA methylation. Thus, the development of new computational and conceptual tools to interpret insect methylation maps is needed. Specifically, we aimed to identify 'highly-methylated regions' (HMR's) in insect genomes using whole genome bisulfite sequencing (WGBS) data. We identified HMR's across several hymenopteran insects and found that similar criteria can be applied to identify most of the methylated CpG's in clusters. We found that most HMR's are located in coding sequences, which supports the idea that methylation is targeted to transcription units in these genomes. In-depth analyses of HMR's in the honey bee further suggest that many HMR's overlap exon/intron boundaries, suggesting a role of DNA methylation on regulation of splicing and/or alternative transcription.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE3

Epigenetic variance as evolutionary strategy for life in an uncertain world

Oana Carja*, Joshua Plotkin [1]

[1]Biology, University of Pennsylvania, Philadelphia, United States

Abstract:

Molecular processes are fundamentally stochastic. Randomness is the rule in transcription, translation, cell-to-cell variation in protein levels, and heterogeneity in interactions. One common assumption is that such phenotypic variation is simply noise, and scientists often appeal to the statistics of large numbers when developing deterministic theories, ignoring any potentially adaptive role of epigenetic stochasticity. Yet evidence is accumulating that epigenetic variance constitutes an evolutionary driving force across diverse biological processes, including the adaptive immune system, the development of cancerous neoplasms, and the persistence of pathogens under drug pressure. All these systems are fundamentally characterized by high levels of environmental change and uncertainty: either persistent, global, temporal fluctuations in selection pressure, or local, micro-environmental and spatially-defined selective forces. In this talk, I will explore the genetic signatures of this commonplace yet unpredictable environmental variation. I will do so by first studying the evolutionary advantage of epigenetic variation as a type of evolutionary bet-hedging that need not confer a direct benefit to a single individual, but can increase the chance of long-term survival of a lineage. I will then discuss a range of population-genetic models that explore how partly heritable phenotypic variability influences evolutionary dynamics of populations. For example, populations that encounter a deleterious environment can sometimes avoid extinction by rapid evolutionary adaptation. I will show that the advantage of such a mutation depends on the degree of phenotypic heritability between generations, called epigenetic memory. Moreover, the probability of population persistence depends non-monotonically on epigenetic memory: some heritability can help avert extinction, but too much heritability removes any benefit of phenotypic plasticity. I will also discuss the implications of these results in the context of therapies designed to eradicate populations of pathogens or aberrant cellular lineages.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-44

Recruitment of histone modifications to assist mRNA dosage maintenance after degeneration of cytosine DNA methylation in model animals

Andrew Chang [1], Ben-Yang Liao [1,*]

[1]National Health Research Institutes, Zhunan, Miaoli County, Taiwan

Abstract: The dosage of mRNA is required to be maintained after gene duplication events. In mammals, this process is achieved with increased cytosine DNA methylation on promoter regions of duplicated genes to suppress transcriptional initiation. However, in many animal lineages, the molecular mechanism of cytosine DNA methylation has degenerated. For such species such as worm (*Caenorhabditis elegans*) and fruit fly (*Drosophila melanogaster*), it is unclear how reduced expression of duplicated genes has been achieved. Here, we hypothesized that epigenetic mechanisms are important for this process, and for organisms whose cytosine DNA methylation apparatus and pattern have degenerated over the course of their evolution, this process is achieved through increased engagement of histone modifications to provide additional levels of epigenetic control to inhibit the expression of duplicated genes, especially genes that are sensitive to changes in dosage. To test our hypothesis, we exploited ENCODE/modENCODE data to examine common and lineage-specific acting regions of several histone marks that are capable of controlling the mRNA abundance of coding genes in worm, fruit fly, and mouse (*Mus musculus*). Our results show that multiple histone modifications are utilized to assist mRNA dosage maintenance after gene duplication events occur in worm, and fly, yet not in mouse. In addition, the histone codes utilized largely differed from one organism to another, suggesting that many of the histone marks that are involved in assisting mRNA dosage maintenance in fly and worm were independently recruited during evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE10

Impact of DNA methylation on the rate and spectrum of mutations in Arabidopsis thaliana

Victoire Baillet [1,*], Leandro Quadrana [1], Mathilde Etcheverry [1], Thomas Bataillon [2], Jean-Marc Aury [3], Patrick Wincker [3], Vincent Colot [1]

[1]Institut de Biologie de l'Ecole Normale Supérieure, PSL Research University, CNRS, INSERM, ENS, Paris, France, [2]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark, [3]Institut de Génomique, Génoscope, CEA, Université d'Evry, CNRS, Evry, France

Abstract: DNA methylation is an epigenetic modification that is pivotal in ensuring proper genome function and integrity, notably through the silencing of transposable elements (TEs). However, as spontaneous deamination of 5-methylcytosine (5mC), which can lead to C>T transitions, is more frequent than that of unmethylated C, DNA methylation is also inherently mutagenic. This higher mutability of 5mC has indeed been proposed to explain the depletion in CpG dinucleotides in mammalian genomes, which are typically methylated at these sites except in so-called CpG islands. Despite this well-characterized effect of DNA methylation, we still lack a comprehensive view of its impact on the whole mutation spectrum in any given organism. Here, we present an in-depth assessment of the mutations that have accumulated in a population of epigenetic recombinant inbred lines (epiRILs) with almost identical genomes but heritable losses of DNA methylation at multiple TE loci in the flowering plant *Arabidopsis thaliana*. As anticipated, the epiRILs have accumulated numerous transposition events, that yet belong to only a small fraction of the TE families defined as mobile at the species level. Similarly, although the rate of C>T transitions is lower in the epiRILs that in classical Mutation Acccumulation (MA) lines, it is at least three times lower than that expected based on the reduced number of 5mC in the epiRILs compared to the MA lines. Along with other hypotheses that are being analyzed, this reduction can be explained by the contribution of transcription-coupled DNA repair. This and other findings will be discussed.

Expanded summary*: DNA methylation is an epigenetic modification that plays important roles in the regulation of gene expression as well as the silencing of transposable elements (TEs). However, DNA methylation is also inherently mutagenic: spontaneous deamination of 5-methylcytosine (5mC) occurs at a higher rate than that of unmethylated C and produces thymine, which can result in a C>T transition if not corrected. This higher mutability of 5mC is indeed thought to underlie the depletion in CpG dinucleotides of mammalian genomes which are typically methylated at CG sites except in so-called CpG islands; and is further illustrated by the large number of disease-associated mutations hotspots found at methylated CpGs in humans.

Despite this well-characterized effect of DNA methylation, we still lack a comprehensive view of its impact on the whole spectrum of mutations in any given organism. Studies to date have been limited in scope and mainly relied on indirect approaches, such as comparisons between methylomes and inter- or intra-species DNA polymorphisms, while one would ideally need to alter DNA methylation patterns in a controlled manner and investigate the mutational consequences genome-wide. This strategy is however difficult to implement, particularly in mammals where a loss of DNA methylation leads to embryonic lethality. In contrast, in the flowering plant *Arabidopsis thaliana*, most DNA methylation mutants are viable and fertile and loss of DNA methylation can in part be stably inherited even in the absence of the inducing mutation.

The population of so-called epigenetic Recombinant Inbred Lines (epiRILs) established in this species by the Colot group (1) provides an ideal experimental system for the investigation of the impact of DNA methylation on the spectrum of spontaneous mutations genome wide. This population derives from a cross between a wild-type individual and a near-isogenic mutant deficient in DNA methylation, and analysis of the epiRILs methylomes revealed that parental differences in DNA methylation (methylated vs unmethylated) are stably inherited over >1000 regions across the genome (2). In addition, the epiRILs resemble by design classical Mutation Accumulation (MA) lines, thus enabling pertinent comparisons between the rate and type of spontaneous mutations generated in the two settings.

In order to describe the extent of TE mobilization upon loss of DNA methylation, whole genome sequencing was carried out for over 100 epiRILs and led to the detection of >1400 new TE insertions, that yet belong to only a small fraction of the TE families defined as mobile at the species level (3; Quadrana, Etcheverry et al, unpublished). Based on these epiRILs sequencing data and on publicly

available MA lines sequences (4), I characterized then compared the full spectrum of single-base substitutions as well as of short indels that have accumulated in these two populations.
The indel spectrum of the epiRILs exhibits an excess of 3 to 5 bp-long insertions which reflects the extensive TE mobilization, as these actually correspond to the genomic footprints that are left upon the excision of a highly mobile DNA transposon. In addition, while the other types of single-base substitution remain unaffected, the rate of C>T transitions is 30-50% lower in the epiRILs. Remarkably, this rate is at least three times lower than that as expected based on the 10-15% reduced number of 5mC in the epiRILs compared to the MA lines. Because the reduction in C>T transitions in the epiRILs almost exclusively concerns TE sequences that are transcriptionally reactivated in this population, these regions may be subjected to transcription-coupled DNA repair, leading to an enhanced repair of deamination-induced mismatches. Current work aims at testing this hypothesis along with several others, and additional analyses are ongoing to characterize the spectrum of large-scale mutations else than transposition events.
This work will provide the first comprehensive assessment at the genome-wide level of the influence of DNA methylation on the complete mutation spectrum in Arabidopsis. This includes obtaining a robust estimate of the rate of methylation-induced C>T transitions and of its determinants, a key metric for many areas of research ranging from comparative genomics to the study of the biology of TEs, which have evolved to GC-poor sequences in many organisms. In addition, in the framework of the genetic assimilation theory proposed by Conrad Waddington, it has been suggested that so-called “epimutations” which would persist long enough to be targeted by natural selection could ultimately become hard-wired in the DNA sequence. In that respect, a proper knowledge of how DNA methylation influences mutation is critical to bring some substance to this hypothesis.
1. Johannes F et al (2009) PLoS Genet 5(6):e1000530.
2. Colome-Tatché M et al (2012) Proc Nat Acad Sci USA 109,16240-5
3. Gilly A (2014) BMC Bioinformatics 15:377.
4. Ossowski S et al (2010) Science 327(5961):92-4

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-423

EPIGENETIC VARIATION BETWEEN URBAN AND RURAL POPULATIONS OF DARWIN'S FINCHES

Sabrina McNew*, Daniel Beck, Ingrid Sadler-Riggleman, Sarah Knutie, Jennifer Koop, Michael Skinner, Dale Clayton

Poster: The molecular basis of evolutionary change is assumed to be genetic variation. However, growing evidence suggests that epigenetic mechanisms, such as DNA methylation, may also be involved in rapid adaptation to new environments. An important first step in evaluating this hypothesis is to test for the presence of epigenetic variation between natural populations living under different environmental conditions. In the current study we explored variation between populations of Darwin's finches, which comprise one of the best-studied examples of adaptive radiation. We tested for morphological, genetic, and epigenetic differences between adjacent "urban" and "rural" populations of each of two species of ground finches, *Geospiza fortis* and *G. fuliginosa*, on Santa Cruz Island in the Galápagos. Using data collected from more than 1000 birds, we found significant morphological differences between populations of *G. fortis*, but not *G. fuliginosa*. We did not find large size copy number variation (CNV) genetic differences between populations of either species. In contrast, we did find dramatic epigenetic differences between the urban and rural populations of both species, based on DNA methylation analysis. We explored genomic features and gene associations of the differentially methylated regions (DMR), as well as their possible functional significance. In summary, our study documents local population epigenetic variation within each of two species of Darwin's finches.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-42

DNA methylation divergence and tissue specialization in the developing mouse placenta

Benjamin Decato [1,*], Jorge Lopez-Tello [2], Amanda Sferruzzi-Perri [2], Andrew Smith [1], Matthew Dean [1]

[1]Molecular & Computational Biology, University of Southern California, Los Angeles, United States, [2]Centre for Trophoblast Research, Dept. of Physiology, Development, and Neuroscience, Cambridge University, Cambridge, United Kingdom

Abstract: The placental epigenome plays a vital role in regulating mammalian growth and development. Aberrations in placental DNA methylation are linked to several disease states, including intrauterine growth restriction and preeclampsia. Studying the evolution and development of the placental epigenome is critical to understanding the origin and progression of such diseases. Although high resolution studies have found substantial variation between placental methylomes of different species, the nature of methylome variation has yet to be characterized within any individual species. We conducted a study of placental DNA methylation at high resolution in multiple strains and closely related species of house mice (*Mus musculus musculus*, *Mus m. domesticus*, and *M. spretus*), across developmental timepoints (embryonic days 15 to 18), and between the labyrinthine transport and junctional endocrine layers. We observed substantial genome-wide methylation heterogeneity in mouse placenta compared to other differentiated tissues. Species-specific methylation profiles were concentrated in retrotransposon subfamilies, specifically RLTR10 and RLTR20 subfamilies. Regulatory regions such as gene promoters and CpG islands displayed cross-species conservation, but showed strong differences between layers and developmental timepoints. Partially methylated domains exist in the mouse placenta and widen during development. Taken together, our results characterize the mouse placental methylome as a highly heterogeneous and deregulated landscape globally, intermixed with actively regulated promoter and retrotransposon sequences.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-51

Human based methyl-capture sequencing for the non-human primate disease models

Jae-Won Huh*, Sang-Je Park [1], Hyeon-Mu Cho [1], Young-Hyun Kim [1], Ja-Rang Lee [1], Se-Hee Choe [1], Sang-Rae Lee [1], Kyu-Tae Chang [1]

[1]National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology, Cheongju, Korea, Republic Of

Abstract: The characterization of genomic or epigenomic variation in human disease models could advance the development of pathophysiological investigation, disease diagnosis, and research of clinical intervention. In the biomedical researches, Macaca fascicularis (Cynomolgus monkey, CM) and Chlorocebus aethiops (African green monkey, AGM) have long been considered important animal model, but non-human primate specific reagents that would evaluate the molecular state of CM and AGM, are racking. Methyl-Capture sequencing (MC-seq) have cost-effectiveness and improved the methylome coverage in large sample sets. Here, we used the human-based MC-seq to assay DNA methylation in 13 CM and 3 AGM blood DNA. Methyl-capture efficiency was associated with sequence identity between the human-based capture probe and each genome sequence of CM and AGM. Approximately 62% and 56% of the human-based capture probes could be reliable mapped to the CM and AGM genome, respectively. To improve the accuracy of human-based methyl-capture for CM and AGM methylome profiling, we redesigned target region, focused on regulatory regions and intragenic regions. Based on our results, we conclude that the human-based MC-seq can be used as the approach for DNA methylome profiling of CM and AGM using the provided filters. Therefore, use of the human-based MC-seq methods provides an attractive, cost-effective approach for the methylome profiling of non-human primate disease models in the single-base resolution level.
Key words: Methyl-capture sequencing; Next generation sequencing; DNA methylation; Non-human primate; Disease model

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-48

Epigenetics of centromere evolution: Hot insights from gibbon LAVA centromeres.

Mariam Okhovat [1,*], Kimberly A. Nevonen [1], Thomas J. Meyer [2], Rachel O'Neill [3], Lucia Carbone [1]

[1]Knight Cardiovascular Institute, Department of Medicine, [2]Division of Bioinformatics and Computational Biology, Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR, [3]Department of Molecular & Cell Biology, University of Connecticut, Storrs, CT, United States

Abstract: Centromeres are crucial for genome stability and chromosome segregation. Despite their conserved function, centromeres are genetically diverse (centromeric paradox). Most centromeres are characterized by repetitive DNA, namely satellite elements and retrotransposons. Centromeric epigenetic modifications are necessary for creating a functional centromere, making it imperative to understand the interplay between epigenetic modifications and centromeric DNA. A striking example of centromere diversity is present in gibbons, which have undergone rapid and drastic karyotype reorganization since diverging from other apes. Particularly, in the *Hoolock* genus, most chromosomes exhibit centromere expansion of a gibbon-specific, non-LTR, retrotransposon, called LAVA (LINE-*Alu*-VNTR-Alu_{like}). While LAVA is found in all other gibbon genera, its centromeric expansion is *Hoolock*-specific. Given its recent centromere reorganization and high inter-chromosomal diversity, the *Holoock* genus is a unique system for studying evolution of centromeres and the function of centromeric repeats. We used chromatin immuneprecipitation-sequencing to examine binding of centromere proteins (CENP) A, B, and C in *Holoock* and other gibbon genera, test the functionality of LAVA expansions and to further identify and compare centromeric repeat content across gibbon species. Our findings will provide invaluable insight into the epigenetics of centromeres, and will bring us closer to solving the centromeric paradox.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-57

Epistasis and Phenotype

Lloyd Brown*

Abstract: There are individuals who, despite having different parents, share identical, or nearly identical, phenotypes. There is currently no explanation for this. It appears to be random. On further investigation, however, it appears to be connected to epistasis. Just as epistasis can be responsible for phenotypes like hair color and eye color, it is also responsible for phenotypes that determine *relationships* between features. These feature-relationships are actually polygenic traits that we have overlooked. Individuals who share these traits look alike despite the fact that they are not related.
One way to look at it is as follows: your overall phenotype is a song. Your inherited traits are the individual notes of the song. The haplogroup in which your genes come from would be the key the song is in. The relationships between the inherited traits would be the *genre* of the song. We have been studying the notes (inherited traits and traits that result from mutation), and all the different keys, but we haven't been paying attention to genre. Songs (phenotypes) in different keys (haplogroups) and with different notes (inherited traits) can still be extremely similar when they are in the same genre, in other words when they are *arranged in the same way*.
The main genres (arrangements) are in the process of being studied and documented, as well as several sub-genres.
Here are some examples that you can see for yourself. [examples of these traits will be presented]

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE14

Using DNA methylation data to test heritability-based predictions of evolutionary models of human aging

Chloe Robins [1,*], Allan McRae [2 3], Joseph Powell [2 3], Howard Wiener [4], Stella Aslibekyan [4], Elizabeth Kennedy [1], Devin Absher [5], Donna Arnett [6], Grant Montgomery [7], Peter Visscher [2 3], David Cutler [1], Karen Conneely [1]

[1]Department of Human Genetics, Emory University, Atlanta, United States, [2]Institue for Molecular Bioscience, [3]Queensland Brain Institute, University of Queensland, Brisbane, Australia, [4]Department of Epidemiology, University of Alabama at Birmingham, Birmingham, [5]Hudson Alpha Institute for Biotechnology, Huntsville, [6]College of Public Health, University of Kentucky, Lexington, United States, [7]QIMR Berghofer Medical Research Institute, Brisbane, Australia

Abstract: The evolutionary theories of mutation accumulation (MA) and disposable soma (DS) provide possible explanations for the existence of human aging. To better understand the relative importance of these theories, we devised a test to identify MA- and DS-consistent sites across the genome using familial DNA methylation (DNAm) data. Two key characteristics of DNAm allowed us to do so. First, DNAm exhibits distinct and widespread changes with age, with numerous age-differentially-methylated (aDM) sites across the genome. Second, many sites show heritable DNAm patterns within families. We extended heritability predictions of MA and DS to DNAm, predicting that MA-consistent aDM sites will show increasing heritability with age, while DS-consistent sites will show the opposite. Variance components models were used to test for changing heritability of methylation with age at 48,601 aDM sites across the genome in 610 individuals from 176 families. 102 sites showed significant MA-consistent increases in heritability with age, while 2,266 showed significant DS-consistent decreases in heritability. These results suggest that both MA and DS play a role in explaining aging and aging-related changes, and that while the majority of DNAm changes observed in aging are consistent with epigenetic drift, targeted changes exist and may mediate effects of age-related genes.

Expanded summary*: The existence of aging remains a fundamental problem in evolutionary biology. Biological aging, or senescence, is the functional decline of an organism with age. This familiar process results in decreased fertility and an increased risk of death, and is associated with major decreases in Darwinian fitness. Based on basic evolutionary theory, such a trait should be opposed by natural selection. Yet aging is universal in many species, including humans. Aging, therefore, poses a paradox. Given the associated decreases in fitness, why has aging not been more effectively selected against? Why is it that humans age?

Evolutionary theories have proposed both adaptive and non-adaptive explanations for the existence of aging. The major theories include: 1) the theory of mutation accumulation (MA), and 2) the theory of antagonistic pleiotropy (AP), with the theory of disposable soma (DS) as a special case. Past empirical tests of these theories have relied heavily on lifespan as the quantitative measure of aging. While lifespan acts as a proxy for an individual's overall rate of senescence, it provides no information on an individual's senescent state or biological age throughout life. Methods of testing evolutionary theories of aging using lifespan as the measure of senescence have provided support for portions of each of theory, suggesting that all three may play some role in the aging process. However, no consensus has been reached on the relative importance of each theory in explaining the existence of senescence. Alternative measures of senescence that are more reflective of aging as a process may provide additional information for such investigations.

DNA methylation (DNAm) has been recently suggested as a biomarker of aging. The methylome has been found to be dynamic throughout life, with several studies reporting thousands of age-differentially-methylated (aDM) sites across the genome. Furthermore, many of these sites show heritable DNAm patterns, where methylation levels are more similar between relatives than unrelated individuals. We hypothesize age-associated DNAm changes to be reflective of the biological aging process, and suggest DNAm data as an innovative measure against which evolutionary theories of aging can be tested.

We use a novel approach to test predictions of evolutionary theories of aging using familial DNAm data. Assuming age-associated DNAm changes are reflective of the aging process, MA suggests that the heritability of DNAm will increase with age, while DS suggests the opposite. We used variance components models to test for age-dependent changes in the heritability of DNAm at aDM CpG sites across the genome. This allowed us to categorize each site as consistent or not consistent with MA or DS, and to better

understand the relative contribution each theory makes in explaining the existence of aging. We observed age-dependent changes in the heritability of methylation that are consistent with both MA and DS, indicating that both theories play a role in explaining human aging and aging-related changes. Furthermore, the number of sites found to have DS-consistent decreasing heritability of methylation was roughly three times the number of sites found to have MA-consistent increasing heritability of methylation. This suggests that the majority of the DNAm changes observed in aging are consistent with DS and epigenetic drift, but some DNAm changes are consistent with MA and may mediate the effects of aging-related genes.
This work utilizes DNAm as a measure of senescence to better clarify the relative importance of the evolutionary theories of MA and DS, and serves to increase our understanding of the DNAm changes we observe in human aging.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE8

Population scale mapping of transposable element variation in Arabidopsis

Tim Stuart [1,*], Steven Eichten [2], Jonathan Cahn [1], Yuliya Karpievitch [1], Justin Borevitz [2], Ryan Lister [1]

[1]Centre for Plant Energy Biology, The University of Western Australia, Perth, [2]Centre for Plant Energy Biology, The Australian National University, Canberra, Australia

Abstract: Transposable elements (TEs) are unique in their ability to deliberately change position within the genome. This process of transposition has the potential to generate much genetic variation over time, and the presence or absence of individual TE insertions represents a major source of genetic variation within populations. We have developed computational methods enabling the identification of these variants from short-read DNA sequence data, and applied these methods to sequencing data for hundreds of wild Arabidopsis accessions. This analysis has revealed an abundance of previously undescribed genetic variation that exists within the species. Furthermore, this variation in TE position provides a genetic basis for epigenetic variation, explaining much of the previously-reported patterns of differential DNA methylation between Arabidopsis accessions. The TE variants identified also have an important influence on the expression of nearby genes, acting both as activators and repressors in a highly insertion-dependent manner. Overall, we demonstrate the considerable variation in TE position that exists in a wild population, and the important effects this variation can have on epigenetic patterns in the genome.

Expanded summary*: Transposable elements (TEs) are mobile genetic elements capable of changing position in the genome. This ability of TEs to move poses a mutagenic potential, and the insertion of TEs into essential regions of the genome may result in deleterious alleles. Subsequently, TE sequences are often silenced through highly complex epigenetic and posttranscriptional pathways. In plants, TEs are often highly methylated in all cytosine DNA sequence contexts, and this pattern of DNA methylation serves as a transcriptional silencing signal to repress movement of the TE. Indeed, inactivation of genes required for the establishment or maintenance of DNA methylation marks can result in the activation and transposition of previously silent TEs. Through the study of many different wild accessions of the model plant Arabidopsis, it has become apparent that these DNA methylation marks that act to silence TEs can vary greatly between accessions in their distribution across the genome. The cause of this observed epigenetic variation, as well as the potential function, has remained unclear.

We developed new computational methods to identify TE presence/absence variation using existing DNA sequencing data. Applying this methodology to data generated by the 1001 genomes project, consisting of hundreds of wild Arabidopsis accessions complete with corresponding DNA methylation data in many cases, we were able to develop a thorough description of TE positional variation in Arabidopsis. We identified over 20,000 TE variants, most of which were present at a low allele frequency.

Using this catalogue of TE variation, we examined patterns of DNA methylation in each accession, and find that most differential DNA methylation between accession can be linked to nearby differential TE insertions, with the insertion of a TE linked with an increase in DNA methylation levels in the surrounding sequence. Furthermore, the deletion of a previously inserted TE was often not associated with a loss of surrounding DNA methylation, indicating that DNA methylation patterns, once established, can be independently maintained over generations in the absence of the genetic trigger. In this way, TEs may leave behind an epigenetic “scar” following their deletion from a particular locus. We argue that TE variation is the genetic basis for much of the

epigenetic variation observed in Arabidopsis, and our work makes an important contribution to our understanding of the genomic impacts of TEs.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-45

The evolution of DNA methylation among baboon species

Tauras Vilgalys [1,*], Clifford Jolly [2 3], Jeff Rogers [4], Jenny Tung [1]

[1]Evolutionary Anthropology, Duke University, Durham, [2]Center for the Study of Human Origins, New York Consortium in Evolutionary Primatology, [3]Anthropology, New York University, New York, [4]Human Genome Sequencing Center and Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, United States

Abstract: Changes in gene regulation have long been thought to play an important role in evolution. However, we know little about the evolutionary forces that shape epigenetic mechanisms of gene regulation across species. Here, we profiled genome-wide DNA methylation patterns in all six extant species of the baboon genus *Papio* (4–14 individuals per species, plus 5 rhesus macaques as an outgroup), a radiation that presents the opportunity to investigate divergence in gene regulation at both shallow and deeper time scales (200,000 – 1.4 million years). We used reduced representation bisulfite sequencing to estimate DNA methylation levels at ~800,000 sites across the genome. In contrast to studies in humans, but similar to studies in great apes, DNA methylation profiles clearly mirror genetic structure. Divergence in DNA methylation between species is correlated with divergence time ($p=3.61x10^{-6}$), but appears to be faster in one of the major baboon clades (*P. cynocephalus, kindae,* and *ursinus*) than the other (*P. anubis, hamadryas,* and *papio,* which occupy the northern part of the baboon range). We detected 150,549 CpG sites with differential methylation between baboons and macaques and 103,636 sites at which variation in DNA methylation is significantly explained by species identity within *Papio* (10% FDR threshold). We also detect species-specific DNA methylation levels that differ from the rest of the genus for each of the six taxa within *Papio* (ranging from ~2,000 – 60,000 sites per species). Initial comparative analysis using Ornstein-Uhlenbeck models suggests that at least some of these changes may be attributable to natural selection.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE5

Pervasive epigenetic effects of euchromatic transposable elements that shape their own population dynamics

Grace Yuh Chwen Lee [1,*], Gary Karpen [1]

[1]Lawrence Berkeley National Laboratory, Berkeley, United States

Abstract: Transposable elements (TEs) are genomic parasites that can increase their copy number at the expense of host fitness. Natural selection against deleterious TE insertions is a potent evolutionary force limiting the selfish increase of TEs, and influencing the population dynamics of TEs. While the deleterious *genetic* effects of TEs have been widely studied, we focus instead on the less explored evolutionary consequences of the *epigenetic* effects of TEs. Euchromatic TEs can be epigenetically silenced via host-directed enrichment of repressive epigenetic marks, which can spread to and influence the function of adjacent sequences. We performed ChIP-Seq with multiple wild-derived *D. melanogaster* and *D. simulans* strains, and provide the first genome-wide quantification of the epigenetic effects of individual TE insertion. Over half of the euchromatic TEs show spread of repressive epigenetic marks to nearby DNA, which can extend up to 20kb. Such effects result in differential epigenetic states and transcript levels of homologous genic alleles, which in return leading to selection against TEs. Interestingly, compared to *D. melanogaster*, TEs in *D. simulans* show stronger epigenetic effects and should be more strongly selected against, which could explain the lower TE content in the species. We provide evidence that this between-species difference in TE's epigenetic effects can be contributed by variation in host genetic factors known to modulate epigenetic silencing. Our study demonstrates that the epigenetic effects of TEs, and host genetic factors modulating such effects, play a critical role in the population dynamics of TEs within and between species.

Expanded summary*: Transposable elements (TEs) are genomic parasites that can increase their copy number at the expense of host fitness. Despite their detrimental effects, TEs constitute appreciable proportions of virtually all eukaryotic genomes. Furthermore, the proportions of host euchromatic genomes that are occupied by TEs are remarkably different between species (i.e. 2.7%–24.9% between 12 *Drosophila* species), and a testable hypothesis for this wide divergence is still lacking. Because of the replicative nature and deleterious fitness effects of TEs, understanding the causes and consequences of their prevalence and variation has remained a central question in evolutionary genomics.

Both theoretical and empirical data have established that selection against the deleterious effects of TE insertions is a potent force limiting the selfish increase in TE copy number, playing an important role in population dynamics of TEs. There is a rich understanding of the deleterious *genetic* effects of TEs mediated by physical disruption of functional elements, such as insertion of TEs into genes. We instead focus on the relatively unexplored evolutionary consequences of the *epigenetic* effects of TEs, and provide evidence for a novel, empirically testable mechanism for the wide divergence in TE content between species.

Active euchromatic TEs can be epigenetically silenced via host-directed enrichment of repressive epigenetic marks. However, as a side effect, these epigenetic marks can spread to and interfere with the function of adjacent sequences. Our epigenomic study identifies extensive epigenetic effects of TEs in *D. melanogaster* and *D. simulans* genomes. Over half of the euchromatic TEs show spread of repressive epigenetic marks to nearby DNA, which can extend up to 20kb (with an average of 4.5kb). Because ~20% of euchromatic genes are within 5 kb from at least one TE insertion in the *D. melanogaster* genome, these results suggest that the functional consequence of TE's epigenetic effects must be extensive. Indeed, the epigenetic effects of TEs result in differential epigenetic states and transcript levels of homologous genic alleles, which in return leading to selection against TE insertions. Importantly, we found substantial variation in the epigenetic effects of TEs between these two closely related species. Based on the molecular understanding of epigenetic regulation, we provide evidence that variation in host genetic factors known to modulate epigenetic silencing could contribute to this between-species difference in TE's epigenetic effects and, ultimately, divergence in TE content.

Our study demonstrates that the epigenetic effects of TEs, and host genetic environment, play a critical role in the evolution of TEs and, more broadly, genomes. Because the molecular mechanisms underlying epigenetic silencing of TEs are similar in plants, insects, and mammals, our work will provide the basis for the future development of unifying models on the evolutionary causes and functional consequences of TE variation across the tree of life.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-49

Developing maps of fitness consequences for plant genomes to uncover functional elements having adaptive value in abiotic stress conditions

Zoe Joly-Lopez [1,*], Adrian E. Platts [1], Adam Siepel [2], Michael D. Purugganan [1]

[1]Biology, New York University, New York, [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, United States

Abstract: As sessile organisms, plants primarily rely on strategies of rapid genomic adaptation to counter environmental threats. Future challenges, exacerbated by climate change, may exceed the adaptive genomic repertoires of many crops, necessitating intervention like targeted mutation of regulatory domains. Thus, we will need to be able to determine the fitness consequences of manipulating such genomic sequences. Computational approaches to mapping fitness can suffer from poor spatial or temporal resolution, but recent approaches may achieve binding-site resolution over lineage-scale intervals. The "fitness consequence" (fitCons) approach is one such example. It classifies a genome relative to shared chromatin states and aggregates these to generate locally corrected but globally inferred levels of constraint using a joint population-comparative model: INSIGHT. This outputs the probability that each nucleotide influences organismal fitness (fitCons scores). To calculate these scores, functional genomic data are used to cluster genomic locations by common fingerprints. Polymorphism and divergence rates are then analyzed using INSIGHT and a fitCons score for each cluster is assigned. The fitCons approach has yet to be explored in plants.

We are developing fitCons maps for plant genomes with the aim of detecting functional elements in noncoding regions that have adaptive value under abiotic stress conditions in the rice crop *Oryza sativa*. To generate fitCons maps, we generate functional genomic tracks from stress-grown rice tissues, which include analyses of gene expression, small RNAs, chromatin states, DNA methylation, and polymerase occupancy.

FitCons maps, combined with traditional genetic approaches, could accelerate discovery of functionally and agronomically important plant genomic elements.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-47

Evolutionary consequences from transposable element mediated epigenetic changes of gene expression in the Asian rice Oryza sativa

Jae Young Choi [1,*], Michael Purugganan [1]

[1]Center for Genomics and Systems Biology, New York University, New York, United States

Abstract: Transposable elements (TEs) are selfish genes that often cause deleterious effects to its host by increasing its own copy number. TEs are believed to exert deleterious effects on the host genome by causing physical disruptions to host genes or from ectopic recombination between TE copies. To counter these detrimental effects, plants such as the rice *Oryza sativa* have evolved epigenetic mechanisms, such as RNA dependent DNA methylation, to silence TE sequences. However, methylation of a TE can be a trade-off as silencing can also spread beyond the TE sequence and affect neighboring host sequences. To examine the genome evolution resulting from the hosts' epigenetic regulation of TEs, we have analyzed transcriptomic and methylomic data from the domesticated Asian rice *O. sativa* ssp. japonica. We then conducted a comparative epigenomic study comparing the japonica TE epigenetic profile with another rice subpopulation indica and its wild progenitor *O. nivara*. Our results indicate that TEs in regions with higher recombination rates and TEs closer to host genes were more highly methylated. However, host genes close to TEs did not show any significant gene expression differences compared to genes further away from TEs. We show that the significant correlation between CHH DNA methylation, which are enriched within heterochromatin-euchromatin boundaries, across TE sequences and distance to nearby host genes may explain the absence of TE-mediated host gene expression changes. Thus our study suggests selection against genome damage is a major driving force of epigenetically silencing TEs with minimal pleiotropic effects on surrounding host gene expressions.

Expanded summary*: Large portion of plant genomes are composed of transposable elements (TEs). In the domesticated Asian rice *Oryza sativa*, 40 – 45% of its ~450Mbp genome is derived from TE related sequences. TEs are generally detrimental to the host by causing physical damage during its integration into the host genome, or by promoting ectopic recombination events. Consequently the host has evolved mechanisms to silence TE activity and for plants these involve RNA dependent DNA methylation of TE sequences. Silencing, however, could be a trade-off as it can also spread beyond the TE sequence and affect host gene or regulatory sequences. Thus, in order to fully understand the TE – host dynamic it is not only important to examine the evolution caused by the physical damage of TEs, but also it is important to examine the evolutionary consequence of epigenetically silencing TEs.

Few studies have integrated the evolutionary genomic, methylomic, and transcriptomic data for investigating TE evolution. Interestingly, these handful of studies have indicated that even within monocots, such as maize and rice, have an opposite relationship between TE methylation level and TE age, suggesting there are species specific evolutionary dynamic between host and its epigenetic regulation of TEs. Further in the case of rice, domestication related bottlenecks could lead to inefficient selection for TE regulation and increased transposition activity. Therefore more studies are necessary to investigate the evolutionary consequence of TE regulation on the host genome. Here, to better understand the evolution of TE regulation I examined patterns of TE methylation and its affect on the host transcription in two subpopulations of *O. sativa* (japonica and indica). I then compared the TE epigenetic profiles of the domesticated *O. sativa* to its wild progenitor *O. nivara*.

Among the different TE families I specifically focused on long terminal repeat (LTR) TEs due to its high abundance across the *Oryza* genome, and its canonical sequence structure makes *de novo* prediction possible. Concordant with previous findings LTR-TEs in japonica were more likely to be methylated if it was longer and younger then average. I further confirmed this in indica and *O. nivara* as well. Further, LTR-TEs located in regions of high recombination rate were more likely to be highly methylated then LTR TEs from low recombination rates. I then compared the reference genomes of japonica and indica to determine orthologous genomic regions and characterize LTR-TEs that were unique or shared between the two genomes. Here, unique LTR-TEs were younger, highly methylated, and interestingly closer to host genes then shared LTR-TEs. These results suggested that highly accessible regions in the genome, such as the euchromatin, might be preferred targets of new LTR-TE insertions. However, because the physical damage caused by TEs are deleterious, new TE insertions are strongly repressed through methylation.

Consistent with this hypothesis LTR-TEs near host genes were more highly methylated then TEs further away from host genes. However, the gene expression levels of host genes near LTR-TEs were not significantly different from genes further away from LTR-TEs. In addition genes near unique LTR-TEs did not have any significant gene expression differences compared to orthologous genes that lacked the LTE-TE. Thus, silencing of TE may be tightly regulated that nearby host gene expression is unaffected, suggesting a minimal trade-off from epigenetically silencing LTR-TEs.
Lastly the results then predicted LTR-TEs would have a higher chance of "surviving" in the host genome if it was further away from host genes. Concordant with this hypothesis there was a positive correlation between LTR-TE age and distance to host gene, but only in the wild progenitor *O. nivara*.
Thus, in this study I have characterized the evolutionary epigenomics of TE regulation in *O. sativa* and found methylation an important mechanism, possibly with minimal pleiotropic effects on surrounding host genic seqences, to silence TE activities from their harmful effects on the host genome. In addition, domestication may have lead to different TE dynamics between the domesticated and wild *Oryza* species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-46

Widespread selection against LTR retrotransposons is balanced by locally adapted transposable element alleles in Arabidopsis thaliana

Michelle Stitzer [1,*], Jeffrey Ross-Ibarra [2]

[1]Center for Population Biology and Department of Plant Sciences, [2]Center for Population Biology, Genome Center, and Department of Plant Sciences, University of California, Davis, Davis, United States

Abstract: Transposable elements (TEs) are exceptional mutagens, and classical theory shows TEs are, on average, slightly deleterious at equilibrium. But averages disguise the variation that exists between TE copies in their impact on the host genome. In the model plant *Arabidopsis thaliana*, selection against TEs operates predominantly through their impact on genes and gene expression, both directly through genetic changes, and when modulated by epigenetic silencing. To investigate variation in the strength of selection acting on TE copies, we characterize 1,635 full-length copies of long terminal repeat retrotransposons in *A. thaliana* reference genomes, and genotype the allele frequency of each TE copy -- as well as 7,500 insertions not present in reference genomes -- in a range-wide sample of 1,135 ecotypes. We relax assumptions of equilibrium by estimating the timing of TE insertion from nucleotide substitutions within the TE, and calculate an expectation of its frequency conditional on age. Although the majority of TE alleles are found at frequencies consistent with widespread negative selection, we identify high frequency alleles consistent with positive selection and a surprising number of polymorphic ancient TEs still likely subject to balancing selection. These non-neutral alleles are overrepresented in associations with cellular phenotypes like local gene expression and methylation patterns at the insertion site as well as flowering time and climatic variables, and many of the TEs identified are from families shown experimentally to transpose in response to abiotic stress. These selective forces reflect the complexity of the struggle between selfish DNA and its host genome.

Expanded summary*: Transposable elements (TEs) are active participants contributing to variation within eukaryotic genomes, but empirical measurements of how natural selection has shaped their frequencies has been recalcitrant due to difficulties in genotyping repetitive sequences. While the mutagenic potential of transposition has been harnessed to great effect in generating phenotypic differences in plant and animal genetic screens, how natural variation in TE content has been shaped by selection has previously relied on investigating specific families of TEs. Although theory suggests TEs at equilibrium should on average be deleterious to the fitness of their host, positive selection on individual TE insertions has also been documented, leading to several hypotheses about the adaptive potential of TE derived variation.

A primary constraint in our understanding of TE polymorphism between individuals is the way in which TEs are annotated in reference genomes. Existing TE annotations in even small genomes like *Arabidopsis thaliana* focus on characterizing repeated sequences. In doing so, these annotations often fail to identify structurally intact transposable elements or TEs that exist in few copies per genome. I have dealt with this limitation by developing a LTR retrotransposon reannotation approach that takes advantage of the structural features of TEs and the biology of transposition to characterize the LTR retrotransposon content in the *A. thaliana* reference genome. Knowing the precise boundaries of the DNA that flanks each TE insertion, and how this delimits the TE itself assists in genotyping TE copies across individuals using short reads. The 1001 genomes

project has released high coverage short read resequencing of 1,135 *Arabidopsis thaliana* ecotypes, with coordinated data on collection site, gene expression, methylation, and flowering time phenotypes for a majority of these ecotypes. I use the resequencing data to genotype LTR retrotransposons, using a combination of methods based on mapping coverage and split read mapping. In addition to calling presence and absence of TEs that exist in the reference genome, I identify 7,500 insertions present at locations not in the reference genome. I validate these TE calls using both the reference assembly of a second *Arabidopsis thaliana* ecotype (Landsberg erecta), and by reassembling TE breakpoints.

Theoretical analysis of transposition-selection equilibrium predicts that most TEs should be rare as weak negative selection limits the proliferation of selfish DNA. I relax these equilibrium assumptions by using sequence divergence in individual TE copies to estimate the age of each allele. Using coalescent simulations accounting for postglacial colonization in *A. thaliana*, I then generate predictions of allele frequency conditional on age. By contrasting empirically measured allele frequencies with those expected from the age of alleles, I identify selection on individual TE copies, even in a background of predominantly negative selection.

In doing so, I identify positively selected TEs and an unexpected number of polymorphic ancient TE copies not in linkage disequilibrium with nearby SNPs, consistent with balancing selection maintaining TE polymorphism. I associate the presence of TE loci with different cellular and plant phenotypes, including gene expression, climate at the collection site, methylation in a region around the insertion, and flowering time. TE families shown experimentally to transpose in response to temperature are enriched within these associations, suggesting that TEs capable of transposition which remain a persistent threat to genome stability are present as a consequence of past selection on TE alleles.

Although TEs make up only a small proportion of the genome in *A. thaliana*, their widespread presence in plants, animals, and fungi suggests that understanding the mechanisms of selection acting on them can assist in moving from associations of genotype and phenotype to characterizing true causal variants. These results highlight the role that LTR retrotransposon variation may contribute to adaptive variation in natural populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE6

Epigenetic mechanisms underlying Paternal Genome Elimination in the scale insect, Planococcus citri

Stevie Anne Bain [1,*], Patrick M Ferree [2], Frances S Turner [3], Laura Ross [1]

[1]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom, [2]W.M. Keck Science Center, Claremont Colleges, Claremont, CA, United States, [3]Edinburgh Genomics, University of Edinburgh, Edinburgh, United Kingdom

Abstract: Paternal genome elimination (PGE) is a genomic imprinting phenomenon found among several insect species, including the citrus mealybugs (*Planococcus citri*). In sons, the paternally inherited set of chromosomes is silenced and then eliminated from the germline. This process is most likely under the control of a maternal factor, since mothers benefit from producing sons that only transmit their genetic material to their offspring. However, it is unclear how the paternal genome in males is targeted for silencing and elimination. The recognition of paternal chromosomes in male *P. citri* appears to be regulated by similar epigenetic machinery to that involved in mammalian silencing and imprinting (histone and DNA methylation). However, the molecular details remain poorly understood and to what extent these mechanisms also play a direct role in transcriptional suppression and germline elimination is inconclusive.

Here we study the parent-of-origin specific epigenetic modifications involved in PGE using molecular and cytogenetic approaches. Whole genome bisulphite sequencing and immunostaining of histone modifications are used to investigate how the parental origins of chromosomes are recognised, allowing specific elimination of the paternal set in sons. Comparing this data with male transcriptome data also gives us insights into how these epigenetic marks are directly involved in gene silencing. Finally, quantitative expression analyses of the genes involved in evolutionarily conserved pathways of transcription repression (DNMT1, SU(VAR)3-9 and HP1) further examines the role these epigenetic modifications play in establishing and maintaining PGE throughout development.

Expanded summary*: Whilst for most organisms it is the case that the copies of a gene they inherit from each parent are interchangeable, some systems defy this rule through the process of genomic imprinting. This process, where expression of one allele is privileged based on its parental origin, is generally considered exclusive to plants and mammals but was actually first discovered in an insect group where silencing and elimination of the entire paternally inherited haploid genome occurs in males. Thus, males can only transmit maternally inherited alleles to their offspring[1]. This is called Paternal Genome Elimination (PGE) and has evolved repeatedly among several insect species, including the citrus mealybug (*Planococcus citri*). Interestingly, PGE appears to be regulated by epigenetic mechanisms similar to those involved in mammalian imprinting (DNA methylation and histone modifications) [2], suggesting the existence of an evolutionarily conserved mechanism for recognizing the parental origin of genes. However, recent work on insects that lack genomic imprinting suggests that there are differences in the epigenetic mechanism between taxonomic groups, for example, DNA methylation is associated with gene silencing in mammals but is either absent or associated with high gene expression in insects[3].

My research focuses on understanding the epigenetic mechanisms that underlie PGE in *Planococcus citri* using molecular and cytological analyses. The main questions and methods of investigation are as follows:

1) ***Are there differences in levels and patterns of DNA methylation associated with chromosomes of maternal and paternal origin in males and females?*** Using whole genome bisulphite sequencing of the offspring of reciprocal crosses between inbred lines with known SNPs.

2) ***How do epigenetic marks affect (parent-of-origin specific) patterns of gene expression in both sexes***? Combining the bisulphite sequencing data with gene expression data from the same experimental crosses.

3) ***When and where are epigenetic marks deposited (in the male or female germline or during fertilization)?*** Immunostaining with antibodies specific to DNA methylation and histone modifications involved in heterochromatinization. Measuring the transcriptional activity of DNA methyltransferases and genes involved in heterochromatinization throughout development in both sexes.

4) ***What is the role of DNA methylation in the silencing of the paternal genome in males?*** Chemically manipulate methylation levels in the male and female germline and study the effects on the phenotype of their (male) offspring.

Understanding how PGE operates, specifically how DNA methylation and histone modifications affect the expression of maternal and paternal alleles will provide general insights into the epigenetic toolbox of life. Parent-of-origin specific gene expression plays an important role in mammalian embryonic development and this research provides the opportunity to develop a tractable invertebrate model system, *Planococcus citri*, for studying parent-of-origin effects that is currently unavailable to the scientific community (since *Drosophila* does not exhibit these effects). This will help us broaden our understanding of how these effects govern development and how they can lead to pathologies when disrupted.

1. Gardner & Ross 2014 Ecology Letters 2. Prantera & Bongiorni, 2012 Genet Res Int 3. Wang *et al*. 2016 *PLoS Biol*

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE1

A possible link between epigenetic regulation of transduplicated gene fragments and microRNA function in maize.

Damon Lisch [1,*], Xinyan Xhang [1]

[1]Botany and Plant Pathology, Purdue University, West Lafayette, United States

Abstract: Epigenetic regulation of transposable elements is an essential feature of all eukaryotic life. In plants, this regulation involves histone modification, DNA methylation, and a variety of small RNAs. One ubiquitous class of small RNAs are those produced by RNA-directed DNA methylation pathway, which uses an RNA-dependent RNA polymerase to produce double-stranded RNA that can be cleaved into small interfering RNAs. In maize, this polymerase is encoded by *Mop1* (*Modifier of paramutation1*). Interestingly, mutations in *Mop1*, like those in the homologous gene in Arabidopsis, have minimal effects on plant morphology. However, after multiple generations in a *mop1* mutant background, we have observed a wide variety of dominant epimutant phenotypes, many of which phenocopy known mutants. One such epimutant exhibits ectopic development of adaxial (inner) leaf tissue on the abaxial (outer) surface of the leaf. We show that this phenotype is caused by ectopic expression of a family of HD-ZIPIII transcription factors, likely due to a reduction in the quantity of the microRNA miR166. Interestingly, we also find evidence for hypomethylation and ectopic expression of a set of Helitron transposons. These are a form of transposon that transposes via a rolling circle mechanism. The vast majority of Helitrons in maize are non-autonomous, meaning that they lack genes encoding proteins that make transposition possible. Instead, many non-autonomous Helitrons carry captured, or transduplicated, fragments of host gene. We find that the ectopically expressed Helitrons are properly spliced and polyadenylated and all carry the target site for miR166. We hypothesized that hypomethylation and transcriptional activation of one or more of these Helitrons in a *mop1* mutant background may result in interference with normal microRNA function. Finally, we speculate as to the degree to which Helitrons may have been recruited as master regulators of particular families of miRNAs. Overall, our data suggests that the presence of vast numbers of transduplicated genic sequences in maize and many other plant species may have had unexpected and consequential effects on gene regulation in plants.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-59

Chromatin state evolution shapes population-level genomic landscapes in Heliconius butterflies

James Lewis [1,*], Robert Reed [1]

[1]Cornell University, Ithaca, United States

Abstract: cis-Regulatory evolution is a key mechanism of biological diversification. Surprisingly little is known, however, about patterns of regulatory variation between intraspecific populations, and the extent to which such variation drives local genomic adaptation. We use functional assays for chromatin accessibility and two histone modifications targeting active promoter and enhancer elements to test the hypothesis that intraspecific genomic divergence is linked to regulatory variation between phenotypically distinct populations of Heliconius butterflies. We show that population-level variability in both chromatin accessibility and cis-regulatory activity, determined via active histone marks, is common within the Heliconius genome. We further demonstrate that differences in cis-regulatory activity between populations do not require associated variation in chromatin accessibility, illustrating that these gene regulatory mechanisms can be evolutionarily decoupled. Importantly, we find that patterns of regulatory variation depart from neutral expectations, suggesting that selection underlies much of the observed regulatory divergence between populations. Supporting this, genomic regions with high Fst are highly enriched for variable regulatory elements, and half of all differentially expressed genes have variable promoter-associated regulatory elements. Our work shows that regulatory elements vary between populations at different functional levels, and that selection on variable elements is a major force underlying genomic divergence within species.

Expanded summary*: This work uses phenotypically divergent subspecies of Heliconius butterflies, well-known for their adaptive mimicry rings, to address three fundamental questions on the nature of cis-regulatory evolution and diversification between both parapatric and allopatric populations of Helionius butterflies: We 1) Investigate the degree to which inter-population variation in cis-regulatory elements drives evolutionary divergence between populations, 2) Assess the biological mechanisms underlying cis-regulatory divergence, and 3) Determine the adaptive significance of cis-regulatory divergence with respect to genome sequence composition (nucleotide divergence) and gene expression profiles (mRNA divergence). This work provides an important benchmark for future investigations into the role of cis-regulatory variation as a driving force behind population diversification and speciation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-54

Single-cell –omics and Exploring the Evolutionary Impact of Germline-Soma Distinctions in Diverse Microbial Eukaryotes

Xyrus Maurer-Alcalá [1 2,*]

[1]Organismic and Evolutionary Biology, University of Massachusetts, Amherst, [2]Biology, Smith College, Northampton, United States

Abstract:

Separate germline and somatic genomes are common across the eukaryotic tree of life, often separated into distinct tissues (e.g. plants, animals and fungi) or as distinct nuclei sharing common cytoplasm (e.g. ciliates and foraminifera). Although in many eukaryotes, the somatic genome contains all the genetic information of the germline (in two copies), this is not true in ciliates whose somatic genomes are highly processed during development and contain only a fraction of the information encoded in the germline. We have used single-cell –omics techniques to explore the evolutionary impact of germline genome architecture and the epigenetically guided processing that occurs in the ciliate *Chilodonella uncinata*. We find that dramatic changes in local composition (GC content) are likely reinforcing the strength of epigenetic mechanisms that delineate somatic and germline genomes and that extensive genome processing (e.g. massive DNA elimination, genome fragmentation and chromosome amplification) is associated with relatively increased rates of protein evolution and provide the basis for the development of expanded gene families.

Expanded summary*:

Separate germline and somatic genomes are common across the eukaryotic tree of life, often separated into distinct tissues (e.g. plants, animals and fungi) or as distinct nuclei sharing common cytoplasm (e.g. ciliates and foraminifera). A substantial amount of discussion exists, outlining the potential evolutionary origins of germline-soma distinctions, which are often tied directly to genome conflict where somatic quiescent germlines reduce the mobility of 'selfish' DNA, limiting their mutagenic power. The aims of my work are two-fold: 1) to address this concept of the evolutionary importance/origins of germline genomes in the context of ciliate (a group of microbial eukaryotes) and 2) developing a workflow that removes strict dependence on cultivation of organisms, which allows for greater comparative work.

From the current state of the work, I have found that emerging single-cell technologies are amazingly apt at circumscribing the traditional laborious tasks of mass cultures, albeit at a cost of assembly size (single-cell techniques produce smaller and more fragmented assemblies). However, the benefits of these techniques vastly outweigh the consequences. As the evolution of germline-soma distinctions are tied to the fundamental question: 'How did sex arise?', the ability to explore germline-soma evolution in an increasingly broad diversity of eukaryotes can provide a better fundamental understanding of evolutionary origin/impact of germline-soma than the current reliance on few 'model' organisms allows.

Incidentally, my work on *Chilodonella uncinata* (a non-model protist) shows how strongly the germline's genome architecture has evolved to better reinforce the epigenetic mechanisms that guide the formation of the somatic genome after sex. Similarly, these data show remarkable similarity across large phylogenetic distances in ciliates (present among clades > 500 MYA apart) and implicate the importance of 'controlled' genome rearrangements (arisen from genome conflict) in developing/expanding ancient gene families.

These data provide better insight into the strong relationship between epigenetically guided developmental processes and the evolutionary significance of the separation of germline and soma. Our data also highlights the emerging importance and feasibility of single-cell –omics techniques in improving our ability to address fundamental questions in evolution and biology by focusing beyond the limited use of 'model' eukaryotes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE7

Genetic and Epigenetic Mechanisms for Gene Expression and Evolutionary Novelty in Plant Polyploids

Z. Jeffrey Chen*

Abstract: Polyploidy is a pervasive evolutionary feature for genomes of some animals and most flowering plants, including many important crop plants such as wheat, cotton, and canola. However, the molecular mechanisms remain largely elusive. Here we report genetic and epigenetic consequences of polyploids in *Arabidopsis* and cotton. We generated a series of resynthesized *Arabidopsis* tetraploids that contain 0–4 copies of *Arabidopsis thaliana* and *Arabidopsis arenosa* genomes and investigated ploidy and hybridity effects on gene expression. Allelic expression can be defined as dosage dependent (expression levels correlate with genome dosages) or otherwise as dosage independent. Here, we show that many dosage-dependent genes contribute to cell cycle, photosynthesis, and metabolism, whereas dosage-independent genes are enriched in biotic and abiotic stress responses. Interestingly, dosage-dependent genes tend to be preserved in ancient biochemical pathways present in both plant and non-plant species, whereas many dosage-independent genes belong to plant-specific pathways. For *A. thaliana* loci, the dosage-dependent alleles are devoid of TEs and tend to correlate with H3K9ac, H3K4me3, and CG DNA methylation, whereas the majority of dosage-independent alleles are enriched with TEs and correspond to H3K27me1, H3K27me3, and CHG (H= A, T, or C) methylation. Furthermore, there is a parent-of-origin effect on nonadditively expressed genes in the reciprocal allotetraploids especially when *A. arenosa* is used as the pollen donor, leading to metabolic and morphological changes. In cotton allotetraploids, histone modifications contribute to genome-wide expression bias of A and D homoeologs. Our experimental evidence suggests that dosage, epigenetic modifications, and cytoplasmic-nuclear interactions shape gene expression diversity in polyploids. During polyploid evolution, genetic and epigenetic mechanisms regulate dosage-dependent expression that can maintain growth and developmental stability, as well as dosage-independent expression that can facilitate functional divergence between homeologs (subfunctionalization and/or neofunctionalization).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-58

Moderate rate of temperature increase leads to transgenerational differences in metabolic and reproduction pathways in a coral reef fish

Moises Bernal [1,*], Jennifer Donelson [2], Heather Veilleux [2], Taewoo Ryu [3], Philip Munday [2], Timothy Ravasi [1]

[1]Biological and Environmental Sciences and Engineering, KAUST , Thuwal , Saudi Arabia, [2]College of Marine and Environmental Sciences, James Cook University, Townsville, Australia, [3]APEC Climate Center , Busan , Korea, Republic Of

Abstract: Human induced climate change is a widespread stressor that can influence the life cycles of a broad array of species. Hence, it is imperative to understand whether species can acclimate to temperature changes forecasted for the upcoming decades. In this study, we used the spiny damselfish (*Acanthochromis polyacanthus*) as a model to understand the effects end-of-century temperatures may have on coral reef fishes. The main objective of this study was to determine if there are differences in acclimation when there is a gradual increase in temperatures across generations. The experimental setup allowed us to compare second-generation individuals from a lineage that reproduced and developed at +1.5C ('intermediate'), versus a similar treatment at +3.0C ('high'). The comparisons also included a group of fishes that developed at +3.0C, but their parents reproduced and developed at +1.5C ('step'). Our results indicate that 'step' individuals cope better than control and 'intermediate' lineages at +3.0C. The assays of gene expression indicated significant differences between second-generation 'step', 'intermediate' and 'high' treatment in genes related to metabolism, reproduction and signal transduction. Overall this study suggests that transgenerational effects differ according to the rate at which individuals are subjected to temperature changes. This system represents an example of how transgenerational adaptations may allow certain species to cope with the impending raise in sea surface temperature.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-53

An Evolutionary Understanding of DNA Methylation Patterns in Nonhuman Primate Skeletal Tissues

Genevieve Housman [1 2,*], Ellen Quillen [3], Anne Stone [1 2]

[1]School of Human Evolution and Social Change, [2]Center for Evolution and Medicine, Arizona State University, Tempe,

[3]Department of Genetics, Texas Biomedical Research Institute, San Antonio, United States

Abstract: Epigenetic mechanisms play crucial roles in the expression of diverse phenotypes within and between species. Specifically, among nonhuman primates (NHPs), DNA methylation patterns have been associated with phylogenetic, behavioral, and disease-related phenotypes. However, this research has primarily focused on soft tissues. This study expands on such exploratory work by assessing the evolutionary relationship of methylation in novel skeletal tissues from NHPs and examining how this variation relates to aspects of bone development and maintenance. Methylation patterns were assessed in femur trabecular bone from baboons (n=74), macaques (n=10), vervets (n=10), chimpanzees (n=4), and marmosets (n=6) using Illumina Infinium MethylationEPIC arrays. Baboons included skeletally healthy adults (n=28), adults with osteoarthritis (n=28), and healthy juveniles (n=18). Within baboons, we identified several differentially methylated positions (DMPs) significantly associated with aging (37.68% of 191,631 sites) and the occurrence of osteoarthritis (0.20% of 191,570 sites). These results reveal the relationship between methylation and bone development and maintenance, and while some of these baboon-specific patterns are conserved with those known in humans, others are not. Among NHPs, out of 39,802 sites examined, we found that 1.63% show species-specific methylation patterns in baboons, 0.65% in macaques, 1.61% in vervets, 7.02% in chimpanzees, and 34.62% in marmosets. Additionally, the global changes in methylation among species reflect known phylogenetic relationships and provide a phylogenetic context for understanding variation known in hominin species. Overall, these findings reveal several DMPs within and among NHPs which begin to inform our evolutionary understanding of this epigenetic mechanism in skeletal tissues within the primate lineage.

Expanded summary*: *Background*

Epigenetic factors play crucial roles in the expression of diverse phenotypes within and between species. The importance of gene regulation for primate phenotypic diversity was originally noted by [1] and has gained credibility as the extent of genetic similarity among phenotypically distinct primate taxa has been clarified. Here we assess this relationship between phenotypic and epigenetic variation to evaluate the magnitude to which epigenetic mechanisms affect phenotypic diversity in nonhuman primates (NHPs).

Within the epigenome, DNA methylation is one mechanism of gene regulation that has the potential to impact phenotypic expression. General changes to mammalian epigenomes have been examined[2], and among NHPs, DNA methylation patterns have been associated with phylogenetic, behavioral, and disease-related phenotypes[3–8]. However, this research has primarily focused on soft tissues. The current study expands on such exploratory work by assessing methylation patterns in novel skeletal tissues from NHPs.

Focusing on skeletal tissues is valuable because inferences from primate skeletal anatomy inform our understanding of primate evolution. Thus, understanding epigenetic contributions to such traits is crucial for proper evaluation of primate skeletal systems. Also, examination of bone pathologies, such as osteoarthritis (OA), have revealed methylation's involvement in bone development and maintenance[9–13]. Methylation likely has similar roles in NHPs; however, this has not been examined. Our research initiates this exportation by assessing the evolutionary relationship of methylation in bone from NHPs and how this variation relates to aspects of bone development and maintenance.

Methods

NHP samples come from captive colonies of chimpanzees (n=4), baboons (n=74), rhesus macaques (n=10), and marmosets (n=6) from the Southwest National Primate Research Center, as well as vervets (n=10) from the Wake Forest/UCLA Vervet Research Colony. Baboons included skeletally healthy adults (n=28), adults with OA (n=28), and healthy juveniles (n=18). Femora were

opportunistically collected at routine necropsy and stored in -20°C freezers. No animals were sacrificed and no living animals were used in this study.

For all NHPs samples, trabecular bone cores were collected from the medial condyle of the right distal femur. Cortical bone was removed, and the remaining bone was pulverized. DNA was extracted[14], and methylation patterns were assessed using Illumina Infinium MethylationEPIC arrays.

Methylation data were normalized[15–17], and probes with poor detection levels were removed from downstream analyses. We additionally filtered out cross-reactive probes, probes containing SNPs at the CpG site, probes detecting SNP information, probes detecting methylation at non-CpG sites, probes targeting sites within the sex chromosomes[15-16,18], probes non-specific to each NHP genome[4,19]. Methylation values for the remaining sites were calculated as the ratio of methylated probe signal intensity to the sum of both methylated and unmethylated probe signal intensities.

To identify sites that were significantly differentially methylated positions (DMPs) for the effect of age, disease status, or taxonomic grouping, we designed and tested generalized linear mixed models (GLMMs) which related the variables of interest to the DNA methylation patterns for each site, while accounting for the effects of additional variables including sex, age, weight, kinship, known batch effects, and unknown latent variables[20–30].

Results

For age, significant DMPs were interrogated from 191,632 sites, and 72,213 DMPs were identified with over half hypermethylated in adults as compared to juveniles. These were associated with several genes, a subset of which overlap with those identified as differentially methylated in human aging studies. In particular, methylation levels of KCNQ1DN have been found to increase in humans with respect to age[31], and this same pattern is observed in baboons.

For disease status, significant DMPs were interrogated from 191,954 sites, and 382 were identified with most hypermethylated in OA as compared to healthy baboons. These were associated with several genes, a subset of which overlap with those identified as differentially methylated in human OA studies. Of those genes that overlapped, some showed identical methylation pattern, while others showed opposing patterns. In particular, HOXA9 and HOXD8 displayed similar OA hypermethylation patterns in baboons and humans[9]. Conversely, LEPR was hypermethylated in OA baboons and hypomethyled in OA humans[32].

For taxonomic groupings, significant DMPs were interrogated from 39,802 sites. The global changes in methylation between species were calculated using Euclidean distances and used to construct species trees. As hypothesized, these topologies reflect known phylogenetic relationships between taxa. Similar pairwise comparisons also identified species-specific DMPs – 650 in baboons, 257 in macaques, 639 in vervets, 2796 in chimpanzees, and 13,778 in marmosets – which are associated with several genes involved in morphological developmental processes, including skin, muscle, brain, and bone development.

Discussion

The primary aim of this research was to assess the evolutionary relationship of methylation in skeletal tissues from NHPs and examine how this variation relates to aspects of bone development and maintenance. Overall, we successfully identified several DMPs within and among NHPs. In particular, the differential methylation observed between adult and juvenile baboons, as well as healthy and OA adult baboons, reveals the relationship between methylation and bone development and maintenance, and while some of patterns are conserved with those known in humans, others are not. Additionally, among NHPs, the differential methylation observed reveals several molecular differences between taxonomic groups which provide possible clues regarding how diverse epigenomes are related to phenotypic variation.

This work is significant because it is an interdisciplinary investigation that merges biological anthropology and epigenetics fields to assess skeletal epigenetics in primates and describe how such features have evolved within this lineage and contributed to phenotypic diversity. Biological anthropologists use variation in primate skeletal anatomy to reconstruct extinct primate species and to support theories on primate phylogenetic and evolutionary history. However, the mechanisms contributing to these hard tissue phenotypes, especially epigenetic factors, are still unclear. The present study is the first to assess methylation variation in hard skeletal tissues from several NHP species, and such explorations are crucial for fully clarifying aspects of primate evolution.

References

1. *Science* 1975, 188:107–16.
2. *Dev Growth Differ* 2010, 52:545–54.
3. *Mol Biol Evol* 2009, 26:1379–89.
4. *PLOS Genet* 2013, 9:e1003763.
5. *Genome Res* 2011, 21:2049–57.
6. *Cell* 2011, 146:1029–41.
7. *PLOS Genet* 2011, 7:e1001316.
8. *Am J Hum Genet* 2012, 91:455–65.
9. *Arthritis Rheum* 2013, 65:197–205.
10. *Gene* 2013, 532:165–72.
11. *PLoS ONE* 2008, 3:e3740.
12. *PLoS ONE* 2014, 9:1–12.
13. *Hum Genet* 2014, 133:1059–73.
14. *Ancient DNA* 2012, 13–9, Humana Press.
15. *Bioinformatics* 2014, 30:1363–69.
16. *bioRxiv* 2016, 65490. doi:10.1101/065490
17. *Nucleic Acids Res* 2013, 41:e90.
18. *Genomics Data* 2016, 9:22–4.
19. *G3 GenesGenomesGenetics* 2014, 4:1227–34.
20. *F1000Research* 2016, 5:1281.
21. *Nucleic Acids Res* 2015, gkv007. doi:10.1093/nar/gkv007
22. *Ann Appl Stat* 2016, 10:946–63.
23. *Nat Methods* 2015, 12:115–21.
24. *Bioinformatics* 2009, 25:765–71.
25. *Stat Sin* 2002, 12:31–46.
26. S*tat Appl Genet Mol Biol* 2004, 3:Article3.
27. *Genome Biol* 2014, 15:R31.
28. *Bioinformatics* 2012, 28:882–3.
29. *Proc Natl Acad Sci* 2008, 105:18718–23.
30. *PLoS Genet* 2007, 3:e161.
31. *Aging* 2011, 3:1018–27.
32. *BMC Musculoskelet Disord* 2015, 16:1–10.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

POA-52

Accurate and consistent DNA methylation age prediction in diverse human populations

Shyamalika Gopalan [1,*], Brenna Henn [1]

[1]Ecology and Evolution, Stony Brook University, Stony Brook, United States

Abstract: Changes in DNA methylation patterns have recently become recognized for their potential to serve as biomarkers of aging in humans. Hundreds of individual CpG sites (cytosine-phosphate-guanine dinucleotides) have been identified that either gain or lose methylation at a steady, predictable rate over organismal age. Measurements of DNA methylation levels from a few to hundreds of CpG sites can be used to construct age prediction models that improve on the accuracy of other biomarkers such as telomere length, amino acid racemization, and mitochondrial DNA deletions. One such algorithm, published in 2013 by Steve Horvath, predicts chronological age from 353 CpG sites and has been shown to be extremely accurate in a variety of tissue types. However, we and others have previously shown that this age calculator produces biased estimates of chronological age in individuals of certain ancestries. The reasons for this bias are not fully understood, but it is clear that this particular age prediction algorithm does not perform consistently across all human populations and genetic backgrounds. Here, we present a new DNA methylation age predictor based on genome-wide DNA methylation data of whole blood and saliva from Europeans, Hispanic-Latinos and Africans. Our model considers a larger set of CpG sites, incorporates meQTLs, and expands the applicability of epigenetic age prediction to diverse ethnicities.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE12

Role of gene body methylation in coral adaptation and acclimatization

Groves Dixon*, Line Bay, Mikhail Matz

Abstract: Gene body methylation is a highly conserved epigenetic mark found in plants and animals, yet its adaptive function remains uncertain. Here we present evidence GBM plays a role in modulating gene expression plasticity in a reef-building coral. Clonal fragments of *Acropora millepora* were reciprocally transplanted between environmentally distinct habitats. After three months, the fragments were assayed for gene expression (Tag-seq) DNA methylation (MBD-seq) to test for associations between GBM, gene expression, and fitness. GBM appears to have a stabilizing effect on transcription. Genes exhibiting between-population differences in GBM showed same-sign differences in transcription, suggesting these genes were stably set do different transcriptional states between the populations. Environmental shifts in GBM were subtle, but positively correlated with transcriptional shifts and with fitness. We suggest that GBM modulates phenotypic plasticity in a positive feedback manner, slowly accumulating in consistently upregulated genes genes, and stabilizing their upregulated state.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Epigenetics and evolution

OM-EE9

Epigenetics and local adaptation in Arabiodopsis lyrata

Nader Aryamanesh [1,*], Weixuan Ning [1], Tuomas Hämälä [1], Outi Savolainen [1]

[1]University of Oulu, Oulu, Finland

Abstract: Epigenetics has become an important element of research in functional genomics, since it is involved in governing the potential to adapt to stress in crops and other plant species through variation in gene regulation. Epigenetic variation may also play a role in local adaptation. In this research, we studied the epigenetic regulation in perennial and outcrossing herb, *Arabidopsis lyrata*. Populations from Germany (Plech) and Norway (high altitude Spiterstulen and lower altitude Lom) were grown and phenotyped in a common garden experiment in two locations with different altitudes (Spiterstulen and Lom) in Norway. Reciprocal transplant experiment showed that the Norwegian populations were locally adapted, with native populations having higher fitness than introduced ones. Whole genome sequencing, whole genome bisulfite sequencing and RNA sequencing was performed to investigate the role of methylation signatures in differential gene expression and local adaptation. Populations showed different methylation pattern and expression profile for different environments. Methylation levels in *Arabidopsis lyrata* were different for CpG, CHG and CHH contexts. The common garden experiment in model perennial plant, Arabidopsis lyrata, made it feasible to dissect short term and long term adaptive responses to the changing environment.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-72

Past, present, and future approaches to the molecular evolution of the first neurons

Benjamin Liebeskind [1,*], Edward Marcotte [2], Richard Aldrich [3]

[1]Institue for Cellular and Molecular Biology, [2]Institute for Cellular and Molecular Biology, [3]Department of Neuroscience, University of Texas at Austin, Austin, United States

Abstract: Nervous systems are one of the most spectacular products of evolution. Their provenance and evolution have been an area of interest and often intense debate for over a hundred years. The genomics era has provided researchers with a new set of tools with which to study the early evolution of neurons, and recent progress on the molecular evolution of the first neurons has been both exciting and frustrating. It has become increasingly obvious that genomic data is often insufficient to reconstruct complex phenotypes in deep evolutionary time. This has been born out in a debate about whether neurons evolved once or twice in early animal evolution. I'll talk about the current state of hypotheses about how and when the first neurons originated and about new methods in proteomics and systems biology of non-model organisms that will pave the way forward.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-92

THE ROLE OF TELOMERES IN THE EVOLUTION OF EXCEPTIONAL LONGEVITY IN BATS

Nicole Foley*, Graham Hughes, Zixia Huang, Michael Clarke, David Jebb, Conor Whelan, Eric Petit, Frédéric Touzalin, Olivier Farcy, Gareth Jones, Roger Ransome, Mary O'Connell, Gerald Kerth, Hugo Rebelo, Luisa Rodrigues, Sébastien Puechmaille, Emma Teeling

Abstract: Ageing is ubiquitous and is associated with an increased risk of suffering from age associated diseases. Evolution has resolved these problems facing our ageing populations in several animal lineages which display negligible senescence. Bats are the longest-lived mammals relative to their body size but it is unknown if telomere shortening, a driving factor of the ageing process, occurs in bats. Uniquely, drawing on >60 years of long-term ringing studies from 4 wild populations of long lived bats (n=448), we show that telomeres shorten with age in *Rhinolophus ferrumequinum* (0-24 years) and *Miniopterus schreibersii* (0-17 years). The genus *Myotis* is enriched with the longest lived species among bats and interestingly, we discovered that there was no significant relationship between telomere length and age in either *M. myotis* (0-6+ years) or *M. bechsteinii* (1-16 years). A comparative analysis of telomerase expression in blood transcriptomes revealed that, like humans, telomerase is not expressed in *M. myotis* blood. Instead, results from tests of positive and divergent selection on 225 telomere maintenance genes derived from 51 eutherian mammals suggest ATM and SETX, which also function to repair and prevent DNA damage, are potential mediators of telomere maintenance in *Myotis*. Furthermore, we show 21 telomere maintenance genes are significantly differentially expressed in *Myotis*, 14 of which directly interact and are enriched for pathways and processes associated with DNA repair. These adaptations detected in long lived *Myotis* bats represent excellent future targets to help us alleviate the human problems associated with ageing.

Expanded summary*: With few exceptions, ageing is ubiquitous and is associated with an increased risk of suffering from age associated diseases. Although mean lifespan in the EU and USA is currently 78 years, and increasing by ~2 years per decade, the age of onset of age related degenerative diseases appears stationary (e.g. 77% of age related cancers occur over 50 years of age). Therefore, we urgently need to better understand the mechanisms of the ageing process with a view to alleviating it in order to improve the future quality of life of our ageing populations. Evolution has resolved these urgent problems facing our ageing populations in a number of animal lineages which display negligible senescence.

Telomeres are tandem TTAGGG repeats which form protective structures at the end of chromosomes. In human somatic cells, telomeres shorten with successive rounds of cell division owing to the 'end replication problem' in which the cellular replicative machinery cannot replicate to the end of the chromosome. As such, telomere shortening with age is recognised as a hallmark of the ageing process and has been observed in a wide diversity of animal populations. Bats are the longest-lived mammals relative to their body size but it is unknown if telomere shortening occurs in bats. Uniquely, drawing on more than 60 years of long-term ringing studies from 4 wild populations of long lived bats (n=448), we show that telomeres shorten with age in *Rhinolophus ferrumequinum* (0 – 24 years) and *Miniopterus schreibersii* (0 – 17 years). The genus *Myotis* is enriched with the longest lived species among bats and interestingly, we discovered that there was no significant relationship between telomere length and age in either *M. myotis* (0 – 6+ years) or *M. bechsteinii* (1 – 16 years).

Telomerase is a reverse transcriptase capable of restoring repeats to telomere ends and is the mediator of telomere maintenance in germ, stem and cancer cells. A comparative analysis of telomerase expression in blood transcriptomes revealed that, like humans, telomerase is not expressed in *M. myotis* blood and as such is unlikely to be involved in telomere maintenance in these species. This result, derived from next-generation-sequencing technologies, contrasts with findings from previous studies which have detected telomerase expression in cell cultures of related species.

To further explore the genetic bases potentially involved in telomere maintenance in *Myotis* we identified 225 target genes using GO terms associated with telomere maintenance and literature searches. Coding sequences from our targeted gene set were mined from 51 publically available eutherian mammal genomes, including 12 bat species, spanning ~98 million years of mammalian evolution. Where possible, data were mined using the RefSeq database. A custom approach incorporating the annotation pipeline Maker, was designed to optimise data mining from poorly annotated genomes. Crucially, this enabled the inclusion of additional genome data

from 5 bat species and 3 mammalian species. Tests of positive and divergent selection were carried out along five branches of our tree, the branch leading to: 1) all bats, 2) the genus *Myotis*, 3) *R. ferrumequinum*, 4) *Miniopterus natalensis* and 5) the naked mole-rat. CodeML selection tests were automated using a custom built workflow called OH-SNAP (**O**ptimised **H**igh-throughput **SN**akemake **A**utomation of **P**AML). Of particular interest our results show that ATM and SETX are evolving under divergent selection in the genus *Myotis*. In addition to the role of these genes in telomere maintenance they also function to repair and prevent DNA damage and as such are identified as potential mediators of telomere maintenance in *Myotis*.

Further support for this result is derived from a multi-tissue comparative transcriptome analysis which showed 21 of these targeted telomere maintenance genes, including ATM and SETX, are significantly differentially expressed in *Myotis* compared to other mammal species. A protein-protein network analysis showed 14 of these differentially expressed genes directly interact and are enriched for KEGG pathways and processes associated with DNA repair. Given that telomerase is present in ~90% of human cancers, therapeutic interventions aimed at enhancing telomere maintenance in the absence of telomerase are desirable. As such, adaptations detected in long lived *Myotis* bats acting on axes intersecting DNA repair and telomere maintenance represent excellent future targets to help us alleviate the human problems associated with ageing.

Of broad importance to the membership of SMBE the research undertaken here for the first time characterises telomere dynamics with age in four populations of long lived bats, representing an unprecedented >60 years of field data. Using a combination of comparative genomics and transcriptomics we begin to unravel the genomic bases potentially underlying telomere maintenance in the exceptionally long lived genus *Myotis* and highlight a role for ATM and SETX. Of broad interest, we describe a novel approach incorporating existing annotation pipelines to better mine unannotated and often poor quality genomes. Furthermore, we describe a custom built work flow, OH-SNAP, which parallelises the widely used single threaded program CodeML from the PAML package, to automate and increase throughput of selection analyses.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

OW-EC6

A novel female reproductive organ - the bursa copulatrix - as a dynamic interface for male-female coevolution

Camille Meslin [1], Melissa Plakke [2], Tamara Cherwin [2], Brandon Small [2], Nathan Morehouse [3], Nathan Clark [4,*]

[1]Institute of Ecology and Environmental Sciences, INRA, Versailles, France, [2]Biological Sciences, University of Pittsburgh, Pittsburgh, [3]Biology, University of Cincinnati, Cincinnati, [4]Computational and Systems Biology, University of Pittsburgh, Pittsburgh, United States

Abstract: During mating, male moths and butterflies introduce a structurally complex ejaculate called a spermatophore that females must digest in order to extract beneficial nutrition and ultimately to mate again. Because the spermatophore is encased in a tough, outer envelope, the female reproductive tract has evolved of a novel organ, the bursa copulatrix, that specifically receives the spermatophore, breaches it and digests the nutrition contained within. This lineage-specific organ provides a unique opportunity to determine the genetic mechanisms by which novel physiological functions arise, and to investigate the evolutionary forces, such as sexual conflict and cooperation, that are thought to have shaped this highly specialized male-female interaction. We have employed transcriptomics and mass spectrometry of 11 male and female tissues along with protein biochemistry and comparative genomics to address these questions in our model, the Cabbage white butterfly (*Pieris rapae*). We determined the male spermatophore is composed of two distinct protein sets that separately form the outer envelope and the inner matrix. These sets are transferred sequentially during mating, with the first forming the envelope via covalent cross-linking of two lepidopteran-specific, proline-rich proteins. In an apparent case of evolutionary convergence, their sequences and amino acid composition resemble unrelated structural proteins such as collagen and spider silk. To counter this substantial barrier to digestion, the female bursa has evolved enlarged musculature that is derived from reproductive tract muscle. Furthermore, many species have evolved a diverse array of hard tools inside the bursa to poke, carve or chew a hole in the spermatophore envelope. Furthermore, the bursa secretes an array of proteases and enzymes to aid in digestion, many of which were coopted from the actual digestive system rather than from the reproductive tract. Using in vitro digestion experiments coupled with mass spectrometry, we demonstrate that these proteases act on a specific set of spermatophore substrates. These proteases are also associated with divergence between subspecies in terms of spermatophore digestion rate in vivo. Finally, we reveal that bursa-specific genes that contribute to these novel functions are relatively rare. Instead, most new functionality is provided by genes coopted from other organ systems. This observation stands in contrast to the male reproductive tract, which is dominated by tissue-specific proteins. However, male and female proteins in this reproductive system do share the common characteristic of evolving at rapid rates. Overall, the bursa-spermatophore interaction presents contrasting mechanisms by which novel functionalities can arise in reproductive tissues. While males produce predominantly tissue-specific proteins to form the spermatophore, the female bursa borrowed and continues to share proteins with non-reproductive organs. These observations suggest that the evolution of complex organ-level phenotypes may initially be enabled by changes in expression patterns that allow expression of existing genes in novel contexts. Subsequent evolution may then lead to gene duplication and further specialization; however, such steps are apparently not required.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-91

The role of adenylate cyclases in the evolution of tanning regulation.

Ellen Quillen [1,*], Anne Sheldrake [1], Jaydee Foster [1], Nina Jablonski [2], Mark Shriver [2]

[1]Genetics, Texas Biomedical Research Institute, San Antonio, [2]Anthropology, Pennsylvania State University, State College, United States

Abstract: Facultative pigmentation, the result of impermanent changes in skin color in response to ultraviolet radiation (UVR) from the sun, is an essential but poorly understood component of the evolution of skin color. Labile facultative pigmentation is the primary interface between the skin and the sun, but the majority of previous research has focused on basal pigmentation levels measured on non-UVR exposed regions of the skin. While basal melanin production is one adaptation to UVR exposure which has evolved several times over human evolution, population-level variation in tanning response and the persistence of newly acquired melanin are convergent mechanisms which could explain variation in facultative pigmentation.

We have recently identified several genes associated with increased tanning response and persistence among 91 Mexican Americans with Indigenous American and European ancestry. Participants received controlled UVR exposures on naïve skin to measure response and persistence following a single exposure at a UVA/UVB ratio equivalent to equatorial sunlight. Measures of melanin content at exposed hand and unexposed underarm were compared to assess the cumulative impact of daily exposure to UVR.

Persistence was calculated as a proportion of facultative melanin (exposed – unexposed) at 28 days/7 days following exposure. Persistence was measured at three levels of UVR exposure. Association with 2950 candidate SNPs within and upstream of KEGG-defined melanogenesis pathway genes was assessed in PLINK while controlling for basal pigmentation and biogeographic ancestry calculated in FRAPPE. All tests were treated as independent, yielding a Bonferoni-corrected α of 5.6×10^{6}. SNPs in eight genes were associated with persistence, including two members of the ubiquitous adenylate cyclase family found on the surface of melanocytes (*ADCY8* and *ADCY9*) and *GNAI1* which regulates adenylate cyclase activity; three growth factors (*FGF10*, *FGF12*, and *MAP2K2*); and transcription factor *CREBBP*.

Associated SNP rs378200 (C > T) is an eQTL for *ADCY9* in unexposed skin based on data from the Genotype-Tissue Expression (GTEx) Consortium ($p = 0.0022$) with the minor allele associated with a reduction in gene expression. Our previously published data on selection at pigmentation loci in Indigenous American populations indicates that selection at *ADCY9* occurred after the split between East Asian and American populations contributing to decreased T allele frequency in the Americas. Identifying and evaluating the evidence of selection on genes that influence constitutive and facultative pigmentation will allow us to better understand this potential source of convergent evolution as early humans faced the old challenge of UVR in new environments.

This work was support by a Post-PhD Research Grant to E. Quillen from The Wenner-Gren Foundation for Anthropological Research.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

OW-EC3

Evolution of alkaloid resistance in poison frogs revealed by comparative phylogenetics

Rebecca Tarvin [1,*], Juan Santos [2], Cecilia Borghese [1], Wiebke Sachs [1 3], Lauren O'Connell [4], Adron Harris [1], Harold Zakon [1], David Cannatella [1]

[1]University of Texas at Austin, Austin, [2]St. John's University, New York, United States, [3]University of Konstanz, Konstanz, Germany, [4]Harvard University, Cambridge, United States

Abstract: Chemically defended organisms must resist their own defenses. Poison frogs (Dendrobatidae) acquire hundreds of chemicals (alkaloids) from their diet as an anti-predator defense. The sheer diversity of alkaloids found in poison frogs (>20 classes, >800 compounds) contrasts substantially with prior studies of toxin resistance, which usually only address organisms that are exposed to one type of chemical. Moreover, alkaloid defenses have evolved at least three times in Dendrobatidae, offering an opportunity to employ comparative approaches to reveal patterns underlying the pathway to alkaloid resistance. Poison frog alkaloids affect important nervous system proteins known as ion channels; resistance to alkaloids is thought to evolve via genetic changes in ion channels where alkaloids bind, i.e., target-site insensitivity. Thus, resistance in poison frogs is a complex trait involving genetic changes in multiple protein families that are targeted by their diverse defensive arsenal. We sequenced several ion channels targeted by poison frog alkaloids in twelve species of poison frogs to identify candidate amino acid sites involved in alkaloid resistance. In voltage-gated sodium channels and nicotinic acetylcholine receptors we found amino acid substitutions that evolved convergently in alkaloid-defended lineages, suggesting they evolved in response to increased exposure to alkaloids. Computational modeling and electrophysiological assays support that these novel substitutions do in fact confer resistance to alkaloids including epibatidine, pumiliotoxins, histrionicotoxins, and batrachotoxins. Moreover, the evolutionary patterns of these substitutions indicate that resistance has evolved multiple times and is an evolutionarily dynamic trait. The level of and potential physiological cost of alkaloid resistance may broadly shape dendrobatid alkaloid profiles, either facilitating or constraining diversification of defenses. Our study advances the understanding of how organisms adapt to wield potent toxins and contributes to the body of knowledge regarding the translation of genotype to complex phenotype.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-93

Unraveling the molecular mechanisms of larval competence, a complex trait that determines dispersal in the sea

Marie Strader [1,*], Phillip Cleves [2], John Pringle [2], Mikhail Matz [1]

[1]Integrative Biology, The University of Texas at Austin, Austin, [2]Stanford University, Palo Alto, United States

Abstract: Dispersal of the majority of large benthic marine invertebrates relies on a planktonic phase of the life cycle dispersed by ocean currents. Patterns and spatial scales of larval dispersal drive biogeographic distributions, genetic connectivity and population and community dynamics. The onset and length of competence, the ability of larvae to metamorphose in response to a specific environmental cue, drives dispersal patterns in the sea. We profiled competence, fluorescence and genome-wide gene expression in embryos and larvae of a reef-building coral *Acropora millepora* throughout thirteen days post-fertilization, tested candidate genes and pathways using a targeted drug screen and developed a method of CRISPR/cas9 knock out on candidate genes. Gene expression associated with competence was positively correlated with transcriptomic response to the natural settlement cue, confirming that mature coral larvae are "primed" for settlement. Rise of competence through development was accompanied by up-regulation of sensory and signal transduction genes such as ion channels, genes involved in neuropeptide signaling, and G-protein coupled receptor (GPCRs). A drug screen targeting components of GPCR signaling pathways confirmed a role in larval settlement behavior and metamorphosis. These results gives insight into the molecular complexity driving this important biological trait and reveals receptors and pathways that, if altered by changing environments, could affect dispersal capabilities of reef-building corals.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POA-413

DEVELOPMENT OF WHOLE GENOME REGULATOR MUTATION LIBRARY IN ESCHERICHIA COLI PROVIDES A PLATFORM BENEFICIAL MUTATIONS TO DIVERSE STRESSES

Alaksh Choudhury [1,*], Zhiwen Wang [2], Joel Kaar [1], Ryan Gill [1]

[1]University Of Colorado, Boulder, Boulder, United States, [2]Tianjin University , Tianjin, China

Poster: Multiple strains of *E coli* have been engineered for bioproduction, bioremediation, drug delivery and biosensing. However, in such modifications, redistribution of cellular resources and non-ideal growth environments create stresses, which compromises cellular fitness. Platforms are required for expedited identification of fitness conferring mutations. Analysis of data from multiple stress tolerance studies, we observed enrichment of mutations in genes involved in regulation of protein expression. However, much of the regulatory mutational space remains unexplored due to limitations in previous approaches. Recently developed Crispr EnAbled Trackable (CREATE) genome Engineering is a tool to enable development of high-throughput site saturation mutagenesis libraries targeting multiple loci across the genome, which can be tracked using unique barcodes. Using, CREATE, we aim to develop libraries with ~178,000 mutations with site saturation mutagenesis of functional sites on regulatory proteins across the *E coli* genome. This work focuses on modifications in the CREATE technology to enable construction of libraries with very high diversity. We demonstrate the use of this library to find mutations that improve fitness for diverse stresses including growth in isobutanol, high salt concentration, non-neutral pH and presence of antibiotics. We also propose to use this library as a platform for studying evolutionary importance of regulatory mutations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-75

Development of craniofacial features in both mice and rats.

Derek Caetano-Anolles*, Sven Künzel, Elke Blohm-Sievers, Abraham Palmer, Oksana Polesskaya, Diethard Tautz

Abstract: Exploring and identifying the genetic components that direct the structure of organisms has remained a major focus of developmental and evolutionary biology. Morphology can evolve very rapidly in populations, which can lead to novel adaptations in response to environmental changes. Here we focus on identifying genes responsible for developing the morphological features of the skull and mandible of the mouse (Mus musculus), while comparing these morphological effects to those found in rat (Rattus norvegicus). Previous Quantitative Trait Locus (QTL) studies have advanced the identification of the genetic basis for craniofacial shape in mice have shown that the phenotype is controlled by genes in several genomic regions. Recently, candidate genes involved in naturally occurring craniofacial shape variation were identified in natural hybrid mice using genome-wide association studies (GWAS). As these were natural hybrids obtained from wild caught mice, the genes identified could therefore be the genes being acted upon by natural selection and affecting craniofacial shape evolution, leading to our use of knockout mice for these candidate genes are being used to confirm their genetic effects on mouse facial development.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

OW-EC1

Evolution of wing development in the Natal long-fingered bat, Miniopterus natalensis

Nicola Illing [1,*], Stephen Schlebusch [1], Mandy Mason [1], Zoe Gill [1], Ash Parker [1], Dorit Hockman [1], Walter Eckalbar [2], Nadav Ahituv [2]

[1]Molecular and Cell Biology, University of Cape Town, Rondebosch, South Africa, [2]Institute of Human Genetics, University of California, San Francisco, San Francisco, United States

Abstract: The bat wing, with its strikingly elongated and webbed fingers contrasting a small, free-toed foot, is used as a model system to study the evolution of morphological variation in vertebrate limbs. We have built up resources for performing developmental and genetic studies on the Natal long-fingered bat, *Miniopterus natalensis*, a populous and gregarious species with a wide distribution in southern and east Africa. This includes a genome assembly (Mnat.v1) of an adult male *M. natalensis*, at 77X coverage that forms the reference for RNA-seq, ChIP-seq (H3K27ac, H3K27me3), ATAC-seq and comparative genomic analyses.

Over 7,000 genes and several lncRNAs, including *Tbx5-as1* and *Hottip*, were differentially expressed between forelimb and hindlimb autopods at three sequential stages of bat development (CS15-CS17); a critical embryonic period when the bat forelimb diverges morphologically from the hindlimb. *In situ* hybridization confirmed the differential expression of the *5' HoxD* and *HoxA* genes, which are known to be essential for patterning the tetrapod autopod. Pathway analysis of RNA-seq data predicted the suppression of the Wnt/b-catenin pathway. This was supported by peanut agglutinin staining, showing larger fields of condensing mesenchymal cells in bat forelimb autopods at the first stages of digit development. In contrast, Wnt-PCP signaling, which maintains the polarity of proliferating chondrocytes in the growth plate, was more active and may set the foundation for extended digit growth at subsequent stages of digit development.

ChIP-seq identified thousands of regions that are differentially modified in forelimb versus hindlimb autopods. These data, combined with comparative genomics was used to pinpoint 2,796 bat-accelerated regions (BARs). Testing five of these BARs for mouse enhancer activity found all of them to be limb enhancer and three of them to show differential enhancer activity compared to the mouse sequence, including a BAR in the HoxD locus. These BARs are candidates for driving the expression changes that led to the evolution of flight in bats 50 million years ago. The challenge is to link these candidate BARs to differentially expressed genes and map out the genetic events that led to the evolution of the bat wing.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-76

Condition-dependent gene expression in the weaponry of Trypoxylus dichotomus

Robert Zinna [1,*], Doug Emlen [2], Laura Lavine [3], Ian Dworkin [4]

[1]Center for Insect Science, University of Arizona, Tucson, [2]Division of Biological Sciences, University of Montana, Missoula, [3]Department of Entomology, Washington State University, Pullman, United States, [4]Department of Biology, McMaster University, Toronto, Canada

Abstract: One of the most dramatic examples of sexual selection are the gigantic weapons used in battles between rival males over access to females. These exaggerated, sexually selected structures tend to be more variable than other body parts and their growth tends to be unusually sensitive to the nutritional state or physiological condition of the individual males who produce and bear them. Despite the importance of weapons to individual fitness, relatively few studies have yet examined the developmental mechanisms responsible for exaggerated growth and heightened condition sensitive expression of sexually selected weapons. Here we use RNA-seq analysis to build on a recent series of studies exploring these mechanisms in the exaggerated weapons of beetles and compare them to non-exaggerated traits, providing an objective screen for differentially transcribed genes associated with a sexually dimorphic and exquisitely condition-sensitive pair of horns in the Japanese rhinoceros beetle. Our results indicate that while genes are differentially expressed according to a trait's degree of condition dependence, relatively few genes change in response to nutrition. On the other hand, sexually dimorphic expression of weaponry involves large-scale changes in gene expression, especially relative to other traits.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-85

Suppression of ancestral mucosal inflammatory response enabled the evolution of extended pregnancy in Eutherian mammals

Arun Chavan [1,*], Oliver Griffith [1], Jamie Maziarz [1], Athanasia Tzika [2], Michel Milinkovitch [2], Gunter Wagner [1]

[1]Yale University, New Haven, CT, United States, [2]University of Geneva, Geneva, Switzerland

Abstract: Pregnancy in Eutherian (Placental) mammals is characterized by its extended duration and ancestrally invasive placentation. Together, these characteristics pose serious immunological challenges to maintenance of pregnancy. The placenta, with its invasive nature can incite an inflammatory reaction in the uterus; and with genetic contribution from the father, can be rejected as 'non-self' by the maternal immune system. Evolving mechanisms to avert these immune complications was central to the evolution of eutherian pregnancy. Here we use comparative transcriptomics of non-gravid and gravid uteri of three eutherians; armadillo (*Dasypus novemcinctus*), tenrec (*Echinops telfairi*), mouse (*Mus musculus*); and one marsupial outgroup; opossum (*Monodelphis domestica*) to study the evolution of immune regulation of eutherian pregnancy. Our data reveal that pregnancy in the viviparous therian ancestor (i.e. the most recent common ancestor of marsupials and eutherians) evoked a combination of type-17 and type-2 immune responses in the uterus. Type-17 is a pro-inflammatory immune response typically deployed in the mucosal surfaces, while type-2 is an anti-inflammatory immune response focused on wound healing. The eutherian lineages sampled are, however, characterized by suppression of type-17 immune response while maintaining an active type-2 response. We conclude that eutherians evolved mechanisms to suppress mucosal type-17 inflammatory response at the fetal-maternal interface, thus enabling maintenance of extended gestation. Using in vitro experiments, we show that this suppression was brought about by uterine decidual stromal cells – an evolutionarily novel cell type in eutherian mammals, and by acquisition of novel functions by progesterone in eutherian lineage.

Expanded summary*: *Question:* Pregnancy in eutherian mammals (i.e. placental mammals) differs in two aspects to the ancestral therian (eutherian and marsupial) or the typical marsupial pregnancy. First, an invasive mode of placentation, where placenta breaches endometrial epithelium and contacts maternal vasculature, evolved in the eutherian stem lineage. Second, eutherians have an extended pregnancy, i.e. pregnancy extending far beyond the estrous cycle duration. Together, these characteristics pose unique immunological challenges to pregnancy, *viz.* (i) uterus is profoundly wounded for an extended period, which can cause inflammation (ii) the invading placenta is semi-allogenic or 'non-self' relative to mother, which can cause immune rejection. In this study we explore how the eutherian immune response evolved to tolerate pregnancy despite these challenges.
Methods: We addressed the above question with a comparative transcriptomic approach. We collected RNA-seq data from the uteri of four therian species in non-pregnant and mid-gestation stages. These include three eutherians: mouse (*Mus musculus*), armadillo (*Dasypus novemcinctus*), tenrec (*Echinops telfairi*); and one marsupial: opossum (*Monodelphis domestica*). These taxa bracket the eutherian phylogeny, and include an outgroup, allowing inference of ancestral eutherian characters. We proposed mechanistic explanations for the observed patterns of evolution of immune regulation, and tested them with in vitro experiments with human and opossum cells.
Results: Inflammation is suppressed in mid-gestation uteri of all species sampled, by a type-2 immune response, an anti-inflammatory response mediated by interleukin-10. The key difference between opossum and eutherians is that a potent type-17 immune response is activated in opossum, while it is completely absent in eutherians. Type-17 is a pro-inflammatory immune response mediated by interleukin-17 and typically deployed in mucosal surfaces. Since type-17 is the default immune response in endometrium (a mucosal surface), and is phylogenetically older than pregnancy, we infer that the type-17 response is the ancestral therian inflammatory response to viviparity. It was ancestrally compensated by simultaneously activating a type-2 response. However, eutherians evolved mechanisms to turn off type-17 inflammation, while leaving the anti-inflammatory type-2 response intact. This resolves the challenge of not only inflammation but also antigenic incompatibility, to some degree, since allo-antigens are not rejected without an accompanying inflammatory response. Using experimental manipulations of human and opossum cells and measurements of immune response in vitro, we identify two mechanisms that contributed to the suppression of mucosal inflammatory response: signaling to immune cells by uterine decidual stromal cells, a novel cell type in eutherian mammals; and acquisition of anti-inflammatory functions by progesterone hormone in eutherian lineage.

Broad Significance: The ability to switch to an anti-inflammatory state after the pro-inflammatory implantation phase is essential for sustaining extended pregnancy in eutherians, and thus understanding the evolution of inflammatory response is central to understanding the evolution of eutherian pregnancy. Evolution of these specialized reproductive strategies may have contributed to the eutherian radiation at Creataceous-Paleogene boundary 65 million years ago. The comparative design of our study complements the extensive biomedical research on reproductive immunology in human and mouse, and uncovers the ancestral role of type-17 inflammation in pregnancy. This study also identifies the role of decidual stromal cells in regulating mucosal inflammation, a potential selective force in the origin of this novel cell-type.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

OW-EC4

The genomic basis of sex change in fish

Erica Todd [1,*], Hui Liu [1], Melissa Slane-Lamm [2], John Godwin [2], Neil Gemmell [1]

[1]Department of Anatomy, University of Otago, Dunedin, New Zealand, [2]Department of Biological Sciences, North Carolina State University, Raleigh, United States

Abstract: Most plants and animals irreversibly differentiate becoming either male or female. In many fishes, sex is not only extremely plastic, but sex change is a natural and adaptive part of the life cycle. Sex change involves coordinated changes in behaviour, neuroendocrinology and gonadal anatomy, typically in response to specific social cues. The molecular basis of this stunning transformation is largely unknown. Using two distantly related wrasses that can be experimentally induced to switch sex (New Zealand spotty *Notolabrus celiodotus*, and bluehead wrasse *Thalassoma bifasciatum*), together with transcriptomic analyses and comparative genomic approaches, we are elucidating the genetic cascade underlying protogynous (female-male) sex reversal. We present RNA-seq differential expression data from gonadal samples representing the complete time-series of sex change. We find that this process involves progressive downregulation of female-specific gene expression, prior to expression profiles becoming increasingly masculinised. We also identify a single female-pathway gene whose immediate downregulation is consistent with the switch initiating protogynous sex change, and develop a hypothetical molecular mechanism that may silence this gene to trigger sex reversal. Many genes classically implicated in vertebrate sexual development are differentially regulated across sex change, suggesting a conserved genetic toolkit has come under the control of environmental cues in sex-changing fish.

Expanded summary*: Sexual fate is no longer seen as an irreversible switch set during early embryonic development, but as an ongoing battle between male and female developmental trajectories. In many fishes, sex is not only extremely plastic, but sex change is a natural and adaptive part of the life cycle. Sex change involves coordinated changes in behaviour, neuroendocrinology and gonadal anatomy, typically in response to specific social cues. However, its molecular basis remains largely unknown.

Using two distantly related wrasses that can be experimentally induced to switch sex (New Zealand spotty *Notolabrus celiodotus*, and bluehead wrasse *Thalassoma bifasciatum*), together with transcriptomic (RNA-seq) analyses and comparative genomic approaches, we are elucidating the genetic cascade underlying protogynous (female-male) sex change. We have extensive data on how behaviour, physiology, and gene expression (brain and gonad) alter during sex change. We find that sex change involves cascaded collapse of female-specific gene expression and replacement by male-specific pathways. We also identify a single female-pathway gene whose immediate downregulation is consistent with the switch initiating protogynous sex change.

We are now investigating the role of epigenetic factors in triggering this switch and regulating sex reversal. Our work couples genome-wide DNA methylation analysis with manipulations of methylation state to establish how methylation alters during sex change and how manipulating methylation affects this process. These experiments will advance our knowledge of how a bi-potential expression network can be differentially regulated to control sexual fate, and will begin to address the boarder question of whether, and how, epigenetic reprogramming is connected with developmental plasticity.

Our work is providing novel insights into the flexibility of sex determination networks, the mechanisms through which developmental plasticity can arise, and the degree to which these mechanisms are conserved evolutionarily. Our findings also have potential practical application in aquaculture and medical research. Many commercially farmed fish change sex and there is considerable interest in controlling sexual fate in aquaculture settings. Sexual gene networks are highly conserved in vertebrates and their non-functioning leads to various disorders of sex development (DSDs) in humans. Sex-changing fish present a unique opportunity to understand the mechanisms through which DSDs arise.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-80

Dissecting Historical Changes of Selective Pressures in the Evolution of Human Pigmentation

Xin Huang [1 2,*], Li Jin [1 3], Yungang He [1]

[1]Chinese Academy of Sciences Key Laboratory of Computational Biology, Chinese Academy of Sciences-Max Planck Society Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Shanghai, [2]Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, [3]State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China

Abstract: Human pigmentation is a highly diverse trait among populations, and has drawn particular attention from both academic and non-academic investigators for thousands of years. To explain the diversity of human pigmentation, researchers have proposed that human pigmentation is adapted for ultraviolet radiation (UVR) and driven by natural selection. Although studies have detected signals of natural selection in several human pigmentation genes, none have quantitatively investigated the historical selective pressures on pigmentation genes during different epochs and thoroughly compared the differences in selective pressures between different populations. In the present study, we developed a new method to dissect historical changes of selective pressures in a multiple population model by summarizing selective pressures on multiple genes. We collected genotypes of 16 critical genes in human pigmentation from 15 public datasets, and obtained data for 3399 individuals of five representative populations from worldwide. Our results suggest (1) that a significant historical increase of selective pressure on light pigmentation shared by all non-Africans at the early stage of the out-of-Africa event (1.78×10^{-2} per generation); (2) that diversifying selection, instead of the relaxation of selective pressures, is the cause of light pigmentation in low UVR areas; (3) and that epistasis plays important roles in the evolution of human pigmentation.

Expanded summary*: Human pigmentation - the color of the skin, hair, and eye - is one of the most diverse traits among populations. Its obvious diversity has attracted particular attention from both academic and non-academic investigators for thousands of years, as noted by Charles Darwin one century ago and as noticed by ancient Egyptians more than 4000 years ago. Why human pigmentation diverges, however, remains a central puzzle in human biology. Human pigmentation may be adapted for UVR and driven by natural selection. Natural selection may favor dark skin for effectively absorbing sunlight, and light skin for efficiently producing vitamin D. Dark skin may protect individuals against sunburn and skin cancer in low latitude areas with high UVR, while light skin may prevent rickets in infants in high latitude areas with low UVR. A better understanding of how natural selection shapes the diversity of human pigmentation could provide relevant and beneficial information for public health.

During the last 10 years, studies have applied methods to detect signals of natural selection in several human pigmentation genes. These genes encode different proteins, such as signal regulators, possible enhancers, important enzymes, and putative exchangers. Although previous studies have been devoted to understanding the evolution of separate pigmentation genes, fewer studies have examined how multiple genes contributed to the evolution of human pigmentation. Moreover, none have quantitatively investigated the historical selective pressures of pigmentation genes during different epochs, and thoroughly compared the differences of selective pressures between different populations.

In the present study, we developed a new method to dissect historical changes of selective pressures for different periods of human evolution. Our results showed not only independent selective pressures in Europeans and Asians, but also a significant historical increase of selective pressure on light pigmentation in all non-Africans at the early stage of the out-of-Africa event. Further, our results excluded the relaxation of selective pressures, and favored diversifying selection as a single explanation for the cause of light pigmentation in low UVR areas, a long-standing puzzle in the evolution of human pigmentation. Finally, our results indicate epistasis plays important roles in the evolution of human pigmentation, partially explaining diversifying selection on human pigmentation among populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-81

Gene network rewiring of convergent evolution of innovative anal fin pigmentation patterns in cichlid fishes

Langyu Gu [1 2,*], Natacha Santo [1], Juan Long [2], Zuzana Musilová [1], Xiaobing Mao [2], Nicolas Boileau [1], Li Li [2], Deshou Wang [2], Walter Salzburger [1]

[1]Basel University, Basel, Switzerland, [2]Southwest University, Chongqing, China

Abstract: The origination and evolution of novelty is the most fascinating question in evolutionary biology. However, how the underlying gene networks are rewired to produce evolutionary novelties is largely unknown. Besides, what are the mechanisms behind the different evolvability of novelties also need to be investigated. The repeated evolution of innovative pigmentation patterns on the anal fin in East African cichlid fish is an ideal model to study these questions. Here we mainly focus on two such patterns: (*i*) egg-spots, i.e. circular markings on the anal fin with different numbers, sizes and colours in the most species-rich lineage, haplochromine lineage. (*ii*) the anal fin blotch with ill-defined boundary found in another independent ectodini lineage. Unlike egg-spots, the blotch shows almost no variation among species. Instead of focusing on individual genes, based on comparative transcriptomic and genomic analysis across the phylogeny, followed by a series of data analyisis such as positive selection detection, transcription factors binding prediction, and functional assays including *in situ* hybridization and transgene, our results proposed that compare to the blotch, egg-spots might evolve a much more independent gene network from the ancestral anal fin, thus providing a clue for its higher evolvability.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-79

Investigating cis-regulatory evolution and sexual selection in Peromyscus mice

Emily Jacobs-Palmer [1], Rachel Agoglia [2,*], Carlo Artieri [3], Kyle Turner [1], Hunter Fraser [3], Hopi Hoekstra [1]

[1]Organismic & Evolutionary Biology, Harvard University, Cambridge, [2]Genetics, [3]Biology, Stanford University, Stanford, United States

Abstract: Promiscuous mating by females of a species is often accompanied by reproductive competition among the males of that species, and has been known to drive precocious male sexual development. In this study, we examine two sister species of deer mice (*Peromyscus maniculatus* and *Peromyscus polionotus*) which exhibit dramatic differences in mating strategy, and consequently display an adaptive divergence in male reproductive timing. We have characterized differences in the timing of male reproductive development in these species and find that in *P. maniculatus*, where females are highly promiscuous, sexual maturity is reached much earlier than in *P. polionotus*, where females mate monogamously. We posit that this divergence has emerged as an adaptive response to differing levels of competition for females, allowing *P. maniculatus* males to compete for females from an early age. To investigate the contribution of *cis*-regulatory evolution on this phenotypic divergence, we performed gene expression profiling on testis tissue from *P. maniculatus* x *P. polionotus* F1 hybrids at ten time points across reproductive development. Analysis of allele specific expression profiles reveals sperm-specific calcium channel subunit, CatsperD, as a compelling candidate gene whose differential expression may contribute to the divergence in reproductive timing. Ultimately, we aim to identify the molecular mechanism by which differential CatsperD expression has evolved, as well as the consequences of this change for male reproductive phenotypes and fitness.

Expanded summary*: Across the animal kingdom, female mate choice drives the evolution of sexually selected phenotypes in males, ranging from unique behaviors to vivid morphological features. In particular, species with promiscuous females – those that mate with multiple male partners – tend to also have reproductive competition among males. In contrast, males of species where females mate monogamously typically lack these adaptive characteristics. In this study, we examine two sister species of deer mice (*Peromyscus maniculatus* and *Peromyscus polionotus*) which exhibit dramatic differences in mating strategy, and consequently display an adaptive divergence in male reproductive timing [1].

Like most mammals, *P. maniculatus* females are highly promiscuous, mating with multiple males within short time frames. In contrast, *P. polionotus* females are strictly monogamous, and males of this species lack certain features of sperm competition (e.g. larger testes and longer sperm) observed in *P. maniculatus* [2]. Much less studied are the differences in reproductive timing between these species. *P. maniculatus* males reach reproductive maturity at a faster rate than *P. polionotus* males, a divergence which we hypothesize has emerged as an adaptive response to differing levels of competition for females, allowing *P. maniculatus* males to compete for females from an early age.

To investigate this hypothesis, we first characterized the differences between the two species in timing of male reproductive development. Through comparing testes size and histology, as well as ability to sire offspring, we found significant differences in age to reproductive maturity between these species, with *P. maniculatus* males reaching all reproductive benchmarks sooner than *P. polionotus.*

To identify candidate genes underlying developmental differences, gene expression profiling is a powerful approach; however the major divergence in the rates of testes development makes between-species comparisons difficult to interpret in this case. We therefore took advantage of these species' ability to form viable F1 hybrids to study allele-specific gene expression, with the goal of identifying signatures of *cis*-regulatory evolution without any confounding effects of changes in developmental rates or cell type abundances. To assess allele-specific expression, we performed RNA-seq on testes samples from hybrid mice at ten time points spanning reproductive development. We found hundreds of genes that are differentially regulated between the two species, and identified a sperm-specific calcium channel subunit, CatsperD, as a compelling candidate gene whose differential expression may

contribute to the divergence in reproductive timing. Ultimately, we aim to identify the molecular mechanism by which differential CatsperD expression has evolved, as well as the consequences of this change for male reproductive phenotypes and fitness.

[1] Dewsbury, D. A. Social dominance, copulatory behavior, and differential reproduction in deer mice (Peromyscus maniculatus). -J. Comp. Physiol. Psychol.95, p. 880-895.
[2] Linzey, A. V., and J. N. Layne. 1969. Comparative morphology of the male reproductive tract in the rodent genus *Peromyscus* (Muridae). American Museum Novitates 2355:1–47.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-84

The tammar wallaby has a eutherian-like placenta that evolved as a molecular tradeoff between lactation and placentation

Michael W. Guernsey [1,*], Edward B. Chuong [2], Guillaume Cornelis [1], Marilyn B. Renfree [3], Julie C. Baker [1]

[1]Department of Genetics, Stanford University School of Medicine, Stanford, CA, [2]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT, United States, [3]School of BioSciences, University of Melbourne, Melbourne, VIC, Australia

Abstract: Therian (live-bearing) mammals nourish their developing young through both placentation and lactation, yet these processes are highly diverse between species. This variation is particularly magnified when comparing marsupial and eutherian lineages, where the marsupial has a morphologically simple placenta and complex lactation while the eutherian has a structurally complex placenta and less dynamic lactation. As both processes are central in nourishing offspring, it has been suggested that marsupials compensate for their simple placenta by expanding their lactation potential. However, despite the lack of morphological complexity, the marsupial placenta functions to maintain pregnancy and embryonic development, indicating that the simple structure contains all of the capabilities of the more complex eutherian organ. In this study we demonstrate that, despite its anatomical simplicity, the tammar wallaby (*Macropus eugenii*) placenta expresses a dynamic molecular program that is highly reminiscent of the eutherian placenta and we provide evidence that gene expression networks that are used to provide nutrients and embryonic growth factors have moved freely between placental and mammary gland cell types throughout the course of morphological evolution. Together, these data provide new molecular insight into the evolution of distinct reproductive strategies employed by therian mammals.

Expanded summary*: Here we utilize transcriptome sequencing in the tammar wallaby to provide evidence that the classification of therian mammals into eutherian (placental) and marsupial (pouch-bearing) based upon having a 'true' placenta is inaccurate. We show that the tammar yolk sac, often argued to be the cognate of the eutherian yolk sac, is surprisingly similar to the mouse and human placenta. Furthermore, immunofluorescence in tammar placenta demonstrates protein localization of genes known to be essential for eutherian placentation to distinct cell layers of the tammar placenta. This striking molecular conservation suggests that although the marsupial placenta is morphologically simple, it represents a 'true' placenta reminiscent of the most complex eutherian forms.
Once we solidify that the tammar wallaby does indeed have a eutherian-like placenta, we examine the evolution of placentation and lactation as strategies for nourishing young. Marsupial biologists have hypothesized that these processes are involved in an evolutionary tradeoff and, as a result, marsupials favored the development of complex lactation while eutherians favored complex placentation to support fetal survival. Interestingly, we find that many of the genetic programs involved in lactation and placentation have been shared, co-opted and exchanged during evolution. These genetic programs have been shifted in both directions: from lactation to placentation and from placentation to lactation. Overall, the recycling of genetic programs to optimize nourishment strategies in each species is a novel finding and surprisingly suggests that placentation and lactation are similar processes.
Successfully producing offspring is at the heart of evolutionary success, yet the mammalian lineage provides a staggering number of ways to achieve this goal. Here we demonstrate that, despite this diversity, conserved molecular programs are utilized to facilitate fetal survival. This example underlies a general hypothesis that molecular programs underlying important organismal functions may move freely between different cell and tissue types throughout the course of morphological evolution. Comparative studies, such as this, enrich our understanding of the tree of life and act as springboard for further investigation into organismal diversity. We also hope that this work will correct widespread misinformation about the most fundamental of biological classifications – that of our own closest relatives.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-86

Observational evidence of directional and stabilizing selection in a contemporary human population

Jaleal Sanjak [1,*], Matthew Robinson [2], Kevin Thornton [1], Peter Visscher [2]

[1]Ecology and Evolutionary Biology, UC Irvine, Irvine, United States, [2]Queensland Brain Institute, Universtiy of Queensland, Brisbane, Australia

Abstract: Evolutionary genetic models are increasingly useful in the study of the genetic architecture of complex traits in humans. However, many fundamental assumptions and parameters of these models remain untested and unmeasured. Medical genetic datasets may allow us to address this problem because they can be used to directly study natural selection in contemporary populations. We leveraged the UK Biobank, a prospective cohort study of over 500,000 phenotyped and genotyped individuals from the United Kingdom, to explore the linear and nonlinear relationships between phenotypes and reproductive success. A number of linear relationships were observed between complex traits and reproductive success using phenotypic data alone, suggestive of directional selection. The evidence also suggests that weak stabilizing selection is common. In general, it appears that stabilizing selection is less intense in humans than in other natural populations, but still may be an important factor shaping the genetic architecture of complex human traits. The evidence of directional selection is further supported by statistically significant estimates of genetic covariance between reproductive success and certain phenotypes via a bivariate linear mixed modeling approach. Finally, we show that finding similar genetic evidence for stabilizing selection is a difficult problem that appears intractable with current datasets and methods.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-82

New regions underlying pigmentation of skin, hair, and iris using quantitative phenotypic traits in individuals of European ancestry

Frida Lona-Durazo [1,*], Melissa Edwards [1], S. Krithika [1], Phuong Le [1], Heather Norton [2], Esteban J. Parra [1]

[1]Anthropology, University of Toronto, Toronto, Canada, [2]Anthropology, University of Cincinnati, Cincinnati, United States

Abstract: A moderate number of genes have been associated with human pigmentary traits (skin, hair and iris color) using genome-wide association studies (GWAS). Most of these GWAS have been carried out in European populations. However, the majority of these efforts have relied on qualitative assessments of eye and hair color, which fail to capture the underlying quantitative distribution of these traits. We performed a GWAS of pigmentary traits in 575 individuals of European ancestry. All traits were evaluated with quantitative methods: skin and hair pigmentation were measured with a reflectometer, and eye color was measured from high-resolution photographs using the CIELab color space. The samples were genotyped with Illumina's Multi-Ethnic Global Array (MEGA), and untyped genotypes were imputed using as a reference the Phase 3 samples of the 1000 Genomes project. We identified signals within well-established genes associated with pigmentation of skin, hair, or iris, such as *IRF4*, *OCA2*, and *HERC2*. We also identified new regions associated with these phenotypes. For skin pigmentation, we identified genome-wide signals ($p<5x10^{-8}$) within or near the genes *WASF1/CDC40* and *EGFR*. For hair pigmentation, genome-wide signals were identified within or near the genes *PRKAA2*, *FHIT*, *MATN2*, *KCNT1* and *DENND5B*. For eye color, we observed a very strong signal within the gene *HERC2*, which has been extensively associated with blue iris color. This region also shows genome-wide significance for central heterochromia. We are currently carrying out replication analyses in independent European samples to confirm the genomic regions identified in our study.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

OW-EC5

Evolution and diversification ef Eye

Atsushi Ogura [1,*]

[1]Nagahama Institute of Bioscience and Technology, Nagahama, Japan

Abstract: The eye is a part of nervous system and connects the internal and external ecologies of organisms by processing visual information. The evolutionary study of eyes aims not only at determining the evolution of the particular organs but also at understanding the role of the organs in the diversification of species. A large number of studies on the developmental biology and molecular biology of eye evolution have revealed that there is a shared developmental process and a common gene regulatory network. Recent work on evolutionary genomics in various types of eyes, together with comparative analyses of gene expression comparison among closely related species, has led to the hypothesis of a dynamic mechanism for the diversification of eyes. Here, we conducted a comprehensive transcriptomic study of developing eyes of Nautilus and pygmy squid. As a result, although most upstream eye development controlling genes were expressed in both species, six3/6 that are required for lens formation in vertebrates was not expressed in Nautilus. Furthermore, many downstream target genes of six3/6 including crystallin genes and other lens protein-related genes were not expressed in Nautilus. As six3/6 and its controlling pathways are widely conserved among molluscs other than Nautilus, the present data suggest that deregulation of the six3/6 pathway led to the pinhole eye evolution in Nautilus. During the transcriptomic study, we also found that there are alternative splicing variants of Pax6 genes in pygmy squid, that gene is known to be important for eye formation. Previous studies have reported that the developmental processes of vertebrate eyes are controlled by four Pax6 splicing variants, each modulating different downstream genes, whereas those of insect eyes are controlled by duplicated Pax6 genes. In the splicing variants, the splicing patterns were produced by the combination of two additional exons to the ortholog and one jettisoned exon containing most of the Homeobox domain (HD). These five variants show spatiotemporally patterns of gene expression during development in the squid. Our study suggests that cephalopods acquired Pax6 splicing variants independent of those in vertebrates and that these variants were similarly utilized in the development of the squid eye.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-69

Novel tissue interactions and the evolution of a new organ, the placenta

Oliver Griffith*, Günter Wagner [1]

[1]Ecology and Evolutionary Biology, Yale University, New Haven, United States

Abstract: Studying the origin and evolution of vertebrate organs is difficult because many have ancient origins and have only evolved once in the history of vertebrates. The placenta is a great model to study the evolution of novel organs because it has evolved repeatedly in many vertebrate clades, it has evolved relatively recently in some lineages, and exists in intermediate forms in extant taxa. By studying a range of vertebrates we investigated the evolution of maternal-fetal interactions in the placenta. In terrestrial vertebrates, placentas form following the interactions of two distinct tissues, the luminal surface of the uterus, and the epithelial surface of an embryonic membrane. Using transcriptomics of the uterus and embryonic membranes from oviparous and viviparous vertebrates we show that the chorioallantoic membrane of amniotes (reptiles, birds, and mammals) was ancestrally an endocrine organ. We then show that the uterus of oviparous reptiles express receptors for the hormones and signalling molecules produced by the chorioallantoic membrane. We argue that the novel apposition of uterine tissues with embryonic membranes is sufficient for the generation of novel maternal-fetal signalling networks. In a second example, we show that key components of implantation in eutherian mammals are the result of an ancestrally inflammatory interaction between the uterus and the apposed embryonic tissue. We argue that the interaction of placental tissues, is not merely a consequence of placenta formation, but that novel interactions form the basis of new placental regulatory networks, functions, and patterning mechanisms.

Expanded summary*: Studying the origin and evolution of vertebrate organs is difficult because many have ancient origins and have only evolved once in the history of vertebrates. The placenta is a great model to study the evolution of novel organs because it has evolved repeatedly in many vertebrate clades, it has evolved relatively recently in some lineages, and exists in intermediate forms in extant taxa.

An insight from studying the evolution of the placenta is that placental traits evolve by utilizing interactions from the novel apposition of maternal and feral tissues. In terrestrial vertebrates, placentas form following the interactions of two distinct tissues, the luminal surface of the uterus, and the epithelial surface of an embryonic membrane.

Using transcriptomics of the embryonic membrane from horse, chicken, the viviparous southern grass skink, the oviparous Australian common garden skink, and both an oviparous and viviparous population of Bougainville's skink, we show that these embryonic membranes are ancestrally endocrine organs. We then show using transcriptomics from oviparous and viviparous skinks that the uterus expresses genes that encode receptors for the produced hormones and signalling molecules are present in the uterus. Using these data, we show that the signalling networks present in viviparous skinks, would also exist in oviparous skinks, if the egg was retained in-utero and the egg shell lost. Therefore, the processes that occur during the evolution of viviparity (egg retention and loss of eggshell) are sufficient for the evolution of maternal-fetal signalling. We argue that the novel apposition of uterine tissues with embryonic membranes is sufficient for the generation of novel maternal-fetal signalling networks.

In a second example, we show that key components of implantation in eutherian mammals are the result of an ancestrally inflammatory reaction of the uterus to the apposed embryonic tissue. One key example of this is the production of prostaglandin E2 (PTGE2). Using transcriptomics and immunohistochemistry, we show that the uterus expresses the first and rate-limiting enzyme in prostaglandin synthesis (PTGS2), but that the final enzyme (PTGES) in PTGE2 synthesis, is expressed only in embryonic placental tissues. Therefore, the synthesis of this enzyme for regulating placental inflammation occurs only when these distinct tissues interact. We argue that the interaction of placental tissues, is not merely a consequence of placenta formation, but that these novel interactions form the basis of new placental regulatory networks, functions, and patterning mechanisms. We propose that the novel apposition of tissues generally is an important process for deriving the unique set of interactions required to support the origin of a new organ.

Significance: We present a new model for understanding the evolution of new organs in vertebrates. Through investigating the evolution of novel placental interactions we show that these interactions arise following the novel apposition of distinct tissue types, a phenomena that could occur elsewhere in an organism following structural rearrangements. This is a significant contribution to this field, because it presents a new model to understand the origin of organs in vertebrates and presents an important experimental system to investigate this model.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

OW-EC2

A gene regulatory network model for the origin of a novel cell type identity

Gunter Wagner [1,*], Cong Liang [2], Arun Chavan [1], Eric Erkenbrack [1], Jamie Maziarz [1]

[1]Ecology and Evolutionary Biology, [2]Computational Biology and Bioinformatics, Yale University, New Haven, United States

Abstract: The origin of novel cell types is a major mode for the evolution of complex body plans in animals. During development cell type identity is established by the activation of core gene regulatory networks (GRN), and thus we suggest that novel cell types arise, in part, through the evolutionary modification of existing core GRNs. Our model system for the origin of a novel cell type is the decidual stromal cell (DSC) of the uterine endometrium of placental mammals. This cell differentiates from endometrial stromal fibroblasts (ESF). Previously we have shown that an outgroup mammal, the opossum, has a cell type homologous to ESF, but these cells do not differentiate into DSC. We reconstructed the core GRN of human DSC via the integration of three distinct datasets: ATAC-seq, RNA-seq and transcription factor binding site matrices. This model is consistent with many known features of human DSC and predicts additional members of the network. We apply the same approach to the endometrial stromal cells of opossum and show that many features of the human DSC core GRN are shared in opossum. For instance, in response to progesterone and cAMP opossum ESF express and post-translationally regulate the FOXO1 protein, which is a key post translationally-regulated transcription factor driving differentiation of DSC in humans. Based on these comparative data we propose a model that predicts which changes in the core GRN likely were necessary for the origin of the DSC during mammalian evolution. We suggest that novel cell types arise through reprogramming of the downstream regulation caused by ancestral signaling systems.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of complex traits

POB-89

Reverse engineering biological complexity through artificial evolution

Yao Yao [1,*], Yves Van de Peer [1 2]

[1]Department of Plant Systems Biology, VIB-UGent Center for Plant Systems Biology, Ghent University, Ghent, Belgium,

[2]Department of Genetics, Genomics Research Institute, Pretoria, South Africa

Abstract: We have developed a bio-inspired robotic simulation framework that mimics the nested architecture of the biological evolutionary process to study biological complexity. Through simulating interactions between different components such as genes and gene regulatory networks, at different levels, from the individual to the population, this framework tries to reproduce the hierarchical and complex structure and context that we observe in biological evolution. Furthermore, our approach suggests a possible model to explain how evolutionary systems might overcome the "costs of complexity" by dynamically and gradually developing different self-organized modular patterns at different scales through the interaction between its different components. Our previous simulation results have shown that such model leads to better adaptability and robustness of the system under a dynamically changing environment while shedding light on the emergence of complex traits. Currently, we are using this platform to examine the effects of polyploidy (whole genome duplication) on the evolutionary process in artificial organisms populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG7

Sequence properties of enhancers are conserved across mammals

Ling Chen [1,*], Alex Fish [2], Tony Capra [1 2]

[1]Biological Sciences, [2]Vanderbilt Genetics Institute, Vanderbilt University, Nashville, United States

Abstract: Gene expression patterns and transcription factor (TF) binding preferences exhibit significant conservation across mammals; however, there is substantial turnover in active enhancers between closely related species. We investigated this seeming contradiction by quantifying the conservation of sequence patterns underlying histone-mark-defined enhancers across six diverse mammals (human, macaque, mouse, dog, cow, and opossum). In each species, we found that machine-learning classifiers based on short DNA motifs could accurately identify adult liver and developing limb enhancers. Applying these classifiers across species, we found that classifiers trained in different species performed nearly as well as classifiers trained on the target species, indicating that the underlying sequence properties of enhancers are largely conserved. We observed similar conservation with enhancers validated in transgenic reporter assays. Supporting the biological relevance of the learned features, the sequence patterns most predictive of enhancers in each species matched a common set of TF motifs that were enriched for expression in relevant tissues. These results suggest that, though the genomic regions with enhancer activity change rapidly between species, short DNA motifs encoding enhancer activity have been maintained across more than 180 million years of mammalian evolution.
To explore the conservation of higher-order regulatory sequence properties, such as cooperative binding of TFs, we are integrating deep neural networks into our cross-species prediction framework. These models are capable of capturing more complex interactions between sequence elements and have shown state-of-art performance in regulatory sequence predictions. We will report preliminary results on the conservation of complex TF interactions found in enhancers across species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-261

Functional dissection of a marine-freshwater adaptive locus that drives lateral line expression in three-spine sticklebacks

Li Ying Tan [1,*], Cecilia Martínez-Rosales [1], William Toubiana [1], Felicity Jones [1]

[1]Friedrich Miescher Laboratory of the Max Planck Society, Tuebingen, Germany

Abstract: Regulatory differences in gene expression are an important source of phenotypic diversity. Therefore, understanding the molecular processes underlying gene expression variation is crucial in achieving a comprehensive picture of adaptive evolution. Three-spine sticklebacks provide a compelling system for studying the molecular basis of adaptation due to the independent colonisation of freshwater habitats by a marine ancestor, resulting in the parallel evolution of freshwater-adapted ecotypes. Moreover, whole-genome sequencing of multiple marine-freshwater population pairs has revealed the predominance of adaptive loci underlying marine-freshwater divergence that potentially involve regulatory changes. Hence, I use transgenic reporter assays in a reverse genetics approach to characterise the regulatory potential of several adaptive loci. Here, I have narrowed a 2500bp adaptive locus into a 500bp intergenic region that drives distinct eGFP activity in the sensory neuromasts of the stickleback lateral line. Neuromasts are mechanoreceptors that allow fish to sense water movement. Sticklebacks inhabiting different environments have been shown to exhibit parallel differences in neuromast number and patterning, which may confer adaptive advantages related to their particular ecology. This small 500bp region also falls within a QTL spanning multiple Mb that was identified in a previous low-resolution mapping study for lateral line variation. Currently, further functional dissection using quantitative reporter assays and genome editing is being carried out to identify the minimal neuromast-specific regulatory element and understand its effects on lateral line phenotype. Ultimately, I aim to elucidate the gene targets of this regulatory element to bridge the gap between genotype and phenotype in adaptively diverging natural populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-262

Adaptive evolution of mental activity-related STX gene in the out-of-Africa migration

Naoko Fujito [1,*], Yoko Satta [1], Masaya Hane [2], Atsushi Matsui [3], Kenta Yashima [1], Ken Kitajima [2], Chihiro Sato [2], Naoyuki Takahata [1], Toshiyuki Hayakawa [4]

[1]School of Advanced Sciences, SOKENDAI(the Graduate University for Advanced Studies), Hayama, [2]Bioscience and Biotechnology Center, Nagoya University, Nagoya , [3]Primate Research Institute, Kyoto University, Kyoto, [4]The Graduate School of Systems Life Sciences, Kyushu University, Hukuoka, Japan

Abstract: It is now reported that a number of genes have undergone adaptive evolution since anatomically modern humans (AMHs) migrated out of Africa. Yet, no such evidence has been found in any gene that is involved in mental activities. It is however conceivable and even likely that AMHs faced mental challenges in the out-of-Africa migration as well as in subsequent new settlements. Here we examine this possibility focusing on a gene that encodes STX, a transferase of polysiallic acids to neural adhesion molecules, and is known to be associated with schizophrenia when overexpressed. There exist three core SNPs that can primarily alter the *STX* promoter activity. These core SNPs define four haplotypes in the current human populations, of which one haplotype, denoted as CGC, is prevalent only in East Asians (though, to a lesser extent in South Asians and Americans as well). We first carried out the promoter assay of the four haplotypes, demonstrating significantly low promoter activity of the CGC. Furthermore, determining 63 haplotype sequences for a world-wide sample, we estimated that the CGC originated ~0.5 MYA and diverged 0.1~0.2 MYA in Africa. We also tested the 1000 genome data in terms of SFS (site frequency spectrum), rEHH and ROH (runs of homozygosity) in a 200 kb region surrounding the core SNPs. All these can be best explained by an ongoing soft sweep of the CGC in East Asian populations, thus providing the first evidence for positive selection on a gene associated with mental activities of AMHs.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-264

Sequence characteristics distinguish enhancers from promoters

Laura Colbran [1,*], Ling Chen [2], John Capra [1]

[1]Vanderbilt Genetics Institute, [2]Biological Sciences, Vanderbilt University, Nashville, United States

Abstract: Promoters and enhancers cooperate to modulate gene expression, thereby playing an important role in all biological processes. While traditionally categorized separately, both primarily serve as binding sites for various transcription factors and might be considered different flavours of the same type of regulatory region. To tease apart potential differences between them, we analyzed sequence characteristics that distinguished 'broad' regions (active across multiple contexts) from 'narrow' regions (active in few) to see whether they were different between promoters and enhancers. In a machine learning framework, 6 bp-long sequences (6-mers) accurately predicted broadly active promoters and enhancers. 6-mers with high GC Content were highly weighted in both promoter and enhancer classifiers. However, in promoters, but not enhancers, CpG dinucleotides were independently associated with regulatory function and wider activity. We tested the ability of classifiers trained on one type of region to distinguish the other, and found that enhancer-trained classifiers were generally able to predict promoters at least as well as they predicted their corresponding enhancer training data. However, the reverse was not true, which suggests that enhancers have sequence characteristics not present in promoters. We tested the most informative 6-mers for similarity to transcription factor (TF) binding motifs, and found that promoter and enhancer classifiers emphasize different families of broadly active TFs. Our analyses suggest that while promoters and enhancers share some key sequence features, enhancers have a layer of complexity that promoter classifiers fail to identify, and highlights a connection between TF motif and expression patterns that differs between the two.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-265

Effects of promoter architecture on gene expression variation due to new mutation

Andrea Hodgins-Davis [1,*], Fabien Duveau [2], Brian Metzger [1], Patricia Wittkopp [1]

[1]EEB, University of Michigan, Ann Arbor, United States, [2]EEB, University of Michigan, Dept of Ecology and Evolutionary Biology, Ann Arbor, United Kingdom

Abstract: Mutations occurring randomly throughout the genome have the potential to have non-random effects on phenotypes based on the regulatory steps required to translate genotypes into phenotypes, *i.e.* the genotype-phenotype map. Differences in the structure of the genotype-phenotype map among phenotypes may bias the phenotypic effects of new mutations, shaping the variation available for evolutionary change. For the phenotype of gene expression, an example of this is the observation that a canonical TATA box motif in a gene's promoter predicts higher mutational variance as well as faster rates of change in experimental evolution and greater polymorphism and divergence. Distinguishing the origin of patterns in phenotypic variation in natural populations thus requires separating the biases introduced by mutation alone from the effects of processes like selection and drift. The Wittkopp lab has previously described the spectrum of new mutations influencing expression of the gene TDH3 and shown that, even within a single gene, local (*cis*) and distal (*trans*) mutations exhibit opposite biases in the magnitude of expression changes they induce: *cis* TDH3 mutations show larger magnitude decreases while trans mutations show larger magnitude increases in expression. However, the extent to which these patterns are generalizable across genes with different promoter architectures and evolutionary histories remains an open question. We will present results from mutagenesis of yeast promoters differing in expression noise, canonical TATA box status, number of predicted regulators, and nucleosome positioning to test hypotheses about the consequences of promoter architecture for the mutational variation that provides the raw material for evolution.

Expanded summary*: Mutations supply the phenotypic variation that guides the course of evolution. If the underlying regulatory networks translating genotypes into phenotypes differ in their susceptibility to new mutations, some regulatory networks may offer more possibilities for variation than others. The Wittkopp lab has previously described the spectrum of new mutations influencing expression of the gene TDH3 and shown that, even within a single gene, local (*cis*) and distal (*trans*) mutations exhibit opposite biases in the magnitude of expression changes they induce: *cis* TDH3 mutations show larger magnitude decreases while trans mutations show larger magnitude increases in expression (Gruber et al 2012, Metzger et al 2016). These patterns are likely to differ substantially among promoters due both to the features of cis regulatory sequence and the structure of upstream regulatory networks that interact with it. While a number of groups are studying the impact of cis changes on gene expression by systematically mutating transcription factor binding sites, promoters, or enhancers in regulatory sequences local to the gene of interest, mutations in trans may be particularly likely to produce changes in gene expression due to a larger potential target size alone. By investigating the impact of new mutations on promoters differing in properties including expression noise, canonical TATA box status, number of predicted regulators, and nucleosome positioning,we test hypotheses about the consequences of promoter architecture for the mutational variation that provides the raw material for evolution. Understanding the extent to which these properties constrain or amplify variation will help us to better interpret the role of forces like natural selection in shaping the patterns of variation observed among and between extant species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG13

How to build a new transcription factor: a novel bat-specific KRAB-transposase fusion gene acts as a transcriptional represssor

Rachel Cosby [1,*], Cedric Feschotte [1], Ellen Pritham [1]

[1]Human Genetics, University of Utah, Salt Lake City, United States

Abstract: Transcription factors (TFs) are fundamental orchestrators of gene regulation. Yet, the mechanisms governing TF birth are poorly understood. Many TFs are derived from coopted transposable elements (TEs), an attractive model considering TEs possess DNA binding domains (DBDs) and disperse their cognate binding sites throughout the genome. Most of these TFs are ancient, however, making it impossible to identify the originating TE and whether it provided the TF's binding sites. In the absence of recent examples, this model has remained largely unexplored. We identify a young (~25my old), bat-specific transposase fusion gene, *KRABINER*, that acts as a transcriptional repressor in reporter assays. *KRABINER* is the product of a complete *mariner* transposase fusion to the KRAB domain of a preexisting KRAB-ZFP. We observe that *KRABINER* is under purifying selection across nine species of bats, and is expressed in at least three species. *KRABINER's* functional domains, KRAB and TE DBD, are complete and required for repression upon binding to intact *mariners*. These data suggest that *KRABINER* is maintained to target *mariners* for silencing. KRAB-mediated repression frequently results in DNA methylation, and *mariners* are methylated in the bat genome. This suggests *KRABINER* might function to repress their activity. Our reporter assays indicate that genes near *mariners* can also be repressed, suggesting that *KRABINER* may regulate host genes. Domain analysis of tetrapod genes indicates that KRAB-transposase fusions occurred no less than 25 times across 20 different lineages. *KRABINER* thus affords an unprecedented opportunity to study a widespread, novel mechanism of genome defense and TF birth.

Expanded summary*: Transcription factors (TFs) are critical orchestrators of gene regulation. Despite this, the age of most TFs makes it difficult to identify the mechanisms driving transcription factor birth. One possible model for TF birth is through cooption of transposable elements (TEs), especially cut-and-paste DNA TEs. In order to mobilize, DNA TEs encode DNA-binding domains (DBDs) that target their own sequences, which are dispersed throughout the genome. Thus, DNA TEs are pre-built sources of DBDs and their cognate binding sequences. If fused to a domain that regulates transcription, cooption of TE DBDs could immediately give rise to a new TF. Indeed, many eukaryotic TFs are derived from TEs; yet these are also ancient, making identification of the originating TE impossible in most cases. Identifying younger TE-derived transcription factors in mammals is especially difficult, given that DNA TEs are extinct in mammalian genomes, except for the vespertilionid (vesper) bats. Without recent examples, this model has remained largely unexplored.

To revisit this model, I examined *de-novo* transcriptomes of vesper bats for transposes fused to other protein domains. In doing so, I identified a novel, lineage-specific transposase fusion gene, *KRABINER*, as a bat-specific isoform of a Kruppel-associated box zinc-finger protein (KRAB-ZFP) gene, *ZNF112*, otherwise conserved in eutherians. *KRABINER* consists of a complete *Mlmar1 mariner* transposase fused in frame to a KRAB domain, resulting from the insertion and subsequent exonization of a *Mlmar1* element into the last intron of the *ZNF112* gene. KRAB-ZFPs are the largest TF family in mammals, and are known to regulate both developmental processes and TE activity via their repressive KRAB domain. RNA-seq and qPCR confirmed both isoforms are expressed in fibroblast cell lines of at least three bats, and the *Mlmar1* insertion responsible for this fusion is orthologous in at least nine species, indicating that *KRABINER* is young (~25 million years old). Selection analysis of these orthologs indicates that *KRABINER* has evolved under purifying selection since its inception ($dN/dS = 0.3$, $p < 0.001$ LRT).

Given the robust repressive activity of the KRAB domain and its fusion to an intact *mariner* DBD, *I hypothesized that KRABINER functions as a transcriptional repressor by binding to Mlmar1 elements in vesper bat genomes.* To test if *KRABINER* is capable of repression, I performed dual-luciferase reporter assays in both human and bat cells. Specifically, I compared luminescence of cells transfected with luciferase driven by either an intact or scrambled *mariner* in the presence of either overexpressed KRABINER or an empty vector. Across three independent experiments in human cells, I determined that KRABINER does repress transcription of luciferase in the presence of an intact *mariner* (MEAN: 2.1% of empty vector; SEM: +/- .007%, n=15) but not in the presence of a scrambled *mariner* (MEAN: 91% of empty vector; SEM: +/- 0.7%, n=15), indicating that its repressive ability is dependent upon

transposase binding. I observed a similar effect in bat cells. Taken together, these data indicate that KRABINER is a transcriptional repressor. Whether or not it binds *Mlmar1* elements in bat genomes remains unknown, and ChIP-seq experiments to test this are currently underway. Nevertheless, the heavily methylated status of *Mlmar1* elements in bat genomes suggests they are repressed by some mechanism. Additionally, if genes near *mariners* are repressed, as the reporter assays suggest, then *KRABINER* could function as a novel TF in bat cells. To test if KRAB-transposase fusions could be a widespread mechanism for TF birth, I examined the domain structure of annotated tetrapod genes. In doing so, I determined that KRAB-transposase fusions occurred no less than 25 times across 20 different lineages. *KRABINER* thus affords an unprecedented opportunity to study a widespread, novel possible mechanism of genome defense and TF birth.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-255

Gene conversion synchronizes expression patterns of paralogous genes

Ajay Chatrath [1], Zhenguo Lin [1,*]

[1]Department of Biology, Saint Louis University, St. Louis, United States

Abstract: Interlocus gene conversion homogenized the sequences of paralogous genes, which led to concerted evolution of gene families. It has not reached a consensus about the rate of interlocus gene conversion and what drives the occurrence of gene conversion. Paralogous genes generated by whole genome duplication events, or ohnologs, have the similar age, which allows us to infer whether gene conversion occurred and duration of concerted evolution. In this study, we developed a progressive method to detect gene conversion and to estimate duration of concerted evolution by analyzing the phylogenetic relationships and sequence divergence of ohnologs from multiple species with different divergence times. Based on analysis of 547 ohnologs from the budding yeast *Saccharomyces cerevisiae* and their orthologs from seven closely related species, we found that up to 32% of ohnologs underwent gene conversion. However, the durations of concerted evolution differ substantially among these ohnologs. The genes with longest durations of concerted evolution are highly enriched in the group of forming macromolecular complexes, such as ribosomal proteins. The similarity of gene expression patterns across many experimental conditions between onologous pairs is positively correlated with duration of concerted evolution, which means that onologs with ongoing gene conversion have most synchronized gene expression patterns. We found that the promoter sequences of these genes were also likely been converted. In summary, our results suggested that gene conversion not only preserves sequencing similarity of paralogous genes but also synchronizes their expression patterns, probably driven by maintaining the stability and normal functions of macromolecular complexes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-260

: Accelerated evolution in ape genomes affects distinct conserved non-coding elements in shared developmental loci

Dennis Kostka [1,*], Alisha Holloway [2], Katherine Pollard [3 4]

[1]Developmental Biology , University of Pittsburgh, Pittsburgh, [2]Phylos Bioscience, Portland, [3]Gladstone Institutes,

[4]Division of Biostatistics, University of Californai San Francisco, San Francisco, United States

Abstract: Accelerated evolution in ape genomes affects distinct conserved non-coding elements in shared developmental loci

Some of the fastest evolving regions of the human genome are conserved non-coding elements with many human-specific DNA substitutions. These Human Accelerated Regions (HARs) are enriched nearby regulatory genes, and many function as developmental enhancers. To investigate if this evolutionary signature is unique to humans, we developed a framework of nested likelihood ratio tests to quantify evidence of accelerated substitutions in conserved genomic elements across multiple lineages. We applied this approach simultaneously to the genomes of five apes: human, chimp, gorilla, orangutan, and gibbon and find roughly similar numbers and genomic distributions of lineage-specific accelerated regions (linARs) in all five apes. In particular, apes share an enrichment of linARs in the distal regions of chromosome arms and in non-coding DNA nearby genes involved in development, especially transcription factors and other regulators. Many developmental loci harbor clusters of distinct linARs from multiple apes, suggesting that accelerated evolution may have affected similar developmental pathways across species. Our statistical tests distinguish between GC-biased and unbiased accelerated substitution rates, allowing us to assess the roles of different evolutionary forces in creating linARs. We find evidence of GC-biased gene conversion in each ape, but signatures consistent with selection are more common in all five lineages. It therefore appears that similar evolutionary processes created independent accelerated regions in the genomes of different apes, and there is a remarkable clustering of these lineage-specific elements in a shared set of developmental loci.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-306

Selection at the pathway level drives the evolution of gene-specific transcriptional noise

Gustavo Barroso [1,*], Julien Dutheil [1]

[1]Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, Ploen, Germany

Abstract: Biochemical reactions within individual cells often result from the interactions of molecules in small numbers, and as such the inherent stochasticity of binding and diffusion processes generate noise along the cascade that leads to the synthesis of a protein from its encoding gene. As a result, isogenic cell populations display phenotypic variability even in homogeneous environments. The extent and consequences of this stochastic gene expression have only recently been assessed on a genome-wide scale, in particular owing to the advent of single cell transcriptomics. The evolutionary forces shaping this stochasticity have yet to be unraveled. We took advantage of three recently published data sets of the single-cell transcriptome of the domestic mouse *Mus musculus* in order to characterize the effect of natural selection on gene-specific transcriptional stochasticity. We showed that noise levels in the mRNA distributions (transcriptional noise) significantly correlate with gene function and gene age, but the main factor that explains observed levels of transcriptional noise is the position of the encoded protein in the biological pathway. Central genes in particular are more deterministically transcribed than the rest. We argue that these results are consistent with models of noise propagation within gene networks. Transcriptional noise is under widespread selection and therefore constitutes an important component of the phenotype. Differences in expression variance – not only in mean expression level – potentially constitute a mechanism of adaptation and should be considered by functional and evolutionary studies of gene expression.

Expanded summary*: Biochemical reactions often result from the interactions of individual molecules in small numbers, and as such the inherent stochasticity of binding and diffusion processes generate noise along the cascade that leads to the synthesis of a protein from its encoding gene. As a result, isogenic cell populations display phenotypic variability even in homogeneous environments. The intensity of this noise is under selection and hence should depend on gene function. We investigated this hypothesis using three publicly available datasets of the mouse single-cell transcriptome.

Currently available measures of expression noise have the drawback of being correlated with the mean expression of genes. This can lead to spurious associations between noise and other functional and/or evolutionary variables if these are correlated with the mean expression level. To sort this out, we fit a linear model on the log-transformed means and variances of all available genes in order to estimate the coefficients *a* and *b* of the resulting power law regression. We used these coefficients to define a new measure of noise as the ratio of the observed variance, and the variance predicted by the linear model:

$F^* = (\sigma^2 / a \cdot \mu^b)$

By construction, F* is independent of the mean expression level, and therefore is a suitable noise metric for investigating the evolutionary forces driving expression noise.

Using the set of genes at both ends of the F* spectrum (i.e., genes with the 10% highest and lowest expression noise) to perform both Gene Ontology and Reactome Pathway enrichment analyses, we showed that the least noisy genes are enriched for functions that relate to the gene expression modulus (i.e., they mostly belong to pathways and ontologies linked to transcription or translation). This result is in line with theoretical predictions that noise should be reduced in top-level genes to avoid deleterious propagation down to other networks. On the other hand, the noisiest genes are not enriched for any ontology or pathway. While some genes can benefit from being stochastically expressed (e.g., to promote phenotypic diversity in the context of fluctuating environmental conditions), they may not have been active in the datasets of unstimulated cells that we analysed.

To better understand the mechanisms of noise propagation, we considered all pathways annotated to the mouse in the Reactome database and computed several network centrality measures for their constituent proteins. At the node level, we showed that the more connections a protein has, the less stochastic is the expression of its encoding gene. Interestingly, proteins with similar F* tend to avoid connecting to each other, which may represent another mechanism to avoid noise propagation inside the network. Likewise, at the whole-network level, pathways whose proteins are more central display lower transcriptional noise, taken as an average among the constitutive genes.

Finally, we used the first axis of the principal component analysis of network centrality measures to define a synthetic variable (SynthNet) that represents the influence of network topology on the level of transcriptional noise of each gene. We combined this

variable with the Ka / Ks ratio of each gene as computed using the human ortholog, as well as the phylostratigraphic gene age, to perform linear model selection with F* as a response variable. The selected model according to Akaike's information criterion has all explanatory variables with a significant effect, but presents SynthNet as the variable that explains most of the variance in F*. We believe that the deleterious effect of noise propagation both within and among pathways is the main driver of natural selection for reduced noise in genes that code for central proteins.

Transcriptional noise is an essential component of an organism's phenotype, in addition to the mean expression level and the actual sequence and structure of the encoded proteins. The study of gene expression must consider changes in noise, in addition to changes in mean expression level, as a putative explanation for adaptation. Our new measure of noise should prove useful to this end. Moreover, the main factor setting the upper threshold in expression noise is the position of proteins inside pathways. That is, central genes are more constrained in order to avoid deleterious noise propagation. Combined, these results should provide insight for future work that further investigates the architecture of noise and how it influences the metabolism of cells under conditions where high noise could be favored (e.g., stimulated immune cells). Given the fundamental nature of biochemical networks, our conclusions should be extendable to a wide range of organisms and therefore have a broad impact in the field.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-259

Inferences of ongoing and resolved sexually antagonistic selection from population genomic data in the flour beetle Tribolium castaneum

Marquerite Herzog [1], Jeff Demuth [1], Amy Dapper [2], Balan Ramesh [1,*]

[1]Biology, Univeristy of Texas at Arlington, Arlington, [2]Genetics, University of Wisconsin, Madison, United States

Abstract: We estimate the degree of ongoing sexually antagonistic selection in flour beetles by comparing sex biased gene expression to differences in allele frequency between the sexes. Organisms with separate sexes often experience selection for different optimal phenotypes in males and females. This sexually antagonistic selection results in sexually dimorphic phenotypes, which are ultimately mediated by sex-biased gene expression. What remains less clear is whether the extensive sex-biased gene expression observed in many organisms is a signature of ongoing sexually antagonistic selection or it reflects antagonism that is already resolved. A recent genome-wide scale analysis in humans and flies reported a "twin peaks pattern" where genes with intermediate levels of sex-biased expression show a stronger signature of ongoing sexual antagonisms (as indicated by higher Fst between sexes) than genes with no sex-bias, or extreme sex-sex bias. Our study takes a similar approach. We use whole genome sequences of individual male and female *Tribolium castaneum* (flour beetle) from 4 populations to calculate Fst between sexes and compare to whole-genome expression data from gonad tissues and whole-body.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-256

Primate iPSCs provide insights into the evolution of adipocyte metabolism

Sasha Makohon-Moore [1,*], Gregory Wray [1]

[1]Biology, Duke University, Durham, United States

Abstract: Humans and our closest living relatives differ in many important traits, including diet and metabolism. Identifying the origin and regulation of these traits is crucial to understanding our own evolutionary history. Further elucidation of the evolutionary basis and molecular underpinnings of many uniquely human traits however, has previously been hampered by the limited access to samples from non-human primates and an inability to perform experiments on these precious samples. The development of induced pluripotent stem cell (iPSC) lines from multiple primate species has recently allowed for renewable access to a variety of cell types, control of genetic effects, reduced environmental variability, and the ability to carry out experimental manipulations. Combining iPSCs with several 'omic approaches provides a toolkit with which we can further understand the evolution of gene expression and regulation in a particular cell type. Adaptive changes in adipocytes, the key component cell of white adipose tissue, were likely crucial during human evolutionary origins. Adipocytes are both the primary energy reserves in the body and a key metabolic regulator furthermore, shifts in metabolic function and diet were of particular importance. We have used adipocytes derived from human and chimpanzee iPSCs to measure genome-wide gene expression and perform regulatory-element sequencing. Together, this allows for the identification of adipocyte-specific differences in molecular, cellular, and metabolic traits. Our results demonstrate the utility of iPSCs for providing insights into the evolution of key metabolic traits that distinguish humans from our closest relatives.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-303

Variation in post-mating transcriptional changes, fecundity and behavior in female D. melanogaster: the roles of female and male genotype

Sofie Delbare [1,*], Clement Chow [2], Mariana Wolfner [1], Andrew Clark [1]

[1]Molecular Biology & Genetics, Cornell University, Ithaca, [2]Human Genetics, University of Utah School of Medicine, Salt Lake City, United States

Abstract: In many organisms, mating induces a multitude of changes in female behavior, physiology and transcriptome. Studies have shown that interactions between female and male genotype lead to variation in post-copulatory phenotypes and reproductive success. In this study, we use the model system *Drosophila melanogaster* to investigate whether such female x male genotype interactions are manifested at the level of the phenotype as well as at the transcriptional level.

To answer these questions, we used *D. melanogaster* inbred lines derived from five geographically dispersed populations. Females from each line were singly mated to males from each of the same five inbred lines. RNA-seq was done on whole mated and virgin females to detect transcriptome changes evident at five to six hours after mating. We fitted linear models to assess whether female genotype, male genotype, or their interaction affected post-mating gene expression changes. In addition, we assayed reproductive output and receptivity.

We find large differences in post-mating phenotypes due to interactions between female and male genotype. On the level of the transcriptome, there are few effects of female and male genotypes separately, which suggests that post-mating gene expression changes are robust across the genetically diverse inbred lines used here. On the other hand, we find that several groups of genes respond strongly to mating only in particular combinations of females and males. These genes are enriched for immune response genes, odorant-binding genes and genes expressed exclusively in the head. In some cases, variation in gene expression is correlated with reproductive output. The transcriptional variation found in specific functional classes of genes might be a read-out of female x male compatibility at a molecular level, and it will be interesting to determine the exact roles these genes play in the female post-mating response.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-257

Genetic control of gene expression in Drosophila: taking eQTL to the next level

Luisa F. Pallares [1,*], Serge Picard [1], Julien F. Ayroles [1]

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, United States

Abstract: Gene expression is not only an intermediate step linking genotypic variation to higher-order phenotypic variation, it is a complex trait that offers unique opportunities to study the regulation of phenotypic variation. To date, key limitations in Drosophila studies have been small sample sizes, the use of inbred lines, and a skewed allele frequency spectrum. This has limited the power to detect genetic associations. We have addressed these limitations by creating a hyper-recombinant outbred population of *Drosophila melanogaster*. This community resource allows us to move away from inbred lines, and assay individual outbred flies. At the same time, we have developed an experimental and analytical approach that allowed us to genotype thousands of flies, individually. We used a genotype-by-sequencing approach to obtain SNP data, and developed a new method (TaqSeq) to obtain gene expression data at a fraction of the cost of existing methods. SNP and RNAseq data was obtained for 5000 individual flies; this resulted in unprecedented statistical power and resolution to detect both cis and trans eQTL, going far beyond previous studies. Our results show the most complete picture of transcriptional regulation in Drosophila, and offer a new perspective on gene regulation in flies by linking genetic to phenotypic variation at the individual level.

Expanded summary*: Phenotypic robustness, understood as the regulation of phenotypic variance is a fundamental characteristic of living organisms. This assures a relative stability of phenotypes even when organisms are exposed to stressful internal or external environments. However, it has been extensively documented that stressful conditions decrease organismal robustness resulting in an increase of phenotypic variance at the population level. Although ubiquitous in biology, the underlying genetic factors driving such shift in variance are poorly understood. My research directly addresses this question: how is phenotypic robustness regulated at the genomic level.

I use flies (*Drosophila melanogaster*) as research organism given the feasibility of obtaining big sample sizes, and the extensive genomic resources available for this system; both are requirements for the success of the project. The phenotype of interest is gene expression. Not only is gene expression an intermediate step linking genotypic variation to higher-order phenotypic variation, but it is, per se, a complex phenotype. It has been documented that the variance around a mean gene expression value is genotype-dependent. Therefore, some individuals have more (or less) potential to deviate from the mean than others. Is this propensity to high (low) variance enhanced under stress? Are the genomic loci controlling variance under standard conditions the same ones controlling this trait under stress? I explore this questions by mapping expressionQTL (eQTL) and variance-eQTL (v-eQTL)[1] in two different environments: standard fly food, and high-sugar diet that represents an stress for the flies.

Previous eQTL studies in Drosophila have used inbred lines, pooled sequencing, and small sample sizes. This have resulted in limited power to detect genetic associations responsible for gene expression variation (eQTL), let alone to map v-eQTLs. We have addressed these limitations by creating a hyper-recombinant outbred population of *Drosophila melanogaster*. This resource allows us to move away from inbred lines, and assay individual outbred flies. To get around sample size limitations, we have made feasible the genotyping and RNA quantification of thousands of individual flies by a fraction of the cost of existing methods. In short, we have developed a genotype-by-sequencing method, and RNAseq protocol based on Tn5 (TagSeq) that allowed us to screen 5000 flies in the control environment and 5000 flies in the high-sugar environment.

The first part of the project, the one I have submitted to present at the SMBE conference, focuses on mapping the genomic loci regulating gene expression (eQTL) in the control environment. Given the differences in the experimental design between this project and previous Drosophila studies, our results have unprecedented statistical power and resolution to detect both cis and trans eQTLs, going far beyond previous studies. We have generated the most complete picture of transcriptional regulation in Drosophila up to date, and offer a new perspective on gene regulation in flies by linking genetic to phenotypic variation at the individual level.

The next step in the project is to identify the genomic loci controlling variance in gene expression (v-eQTL) in the control environment. The eQTL and v-eQTL maps will be defined for the high-sugar environment. The comparison between control and stress conditions is the ultimate goal of this research. It will allow us to reveal cryptic genetic variants only relevant under stress, and to explore how stress disturbs the co-expression networks present in normal conditions. The analysis of this data will reveal the

genomic loci associated with variation in transcriptional robustness between individuals, and will ultimately offer a deeper evolutionary and medical understanding of phenotypic robustness.

[1] v-eQTL are loci whose allelic state predicts the amount of variability around the expected mean.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-254

Re-shaping the map of adaptation in great apes.

Alejandro Berrio [1,*], Lee Elizabeth Edsall [1], Gregory Crawford [1], Gregory Wray [1]

[1]Duke University, Durham, United States

Abstract: Adaptive changes in *cis*-regulatory elements are an important component of evolution by natural selection. Several statistical methods (e.g. branch-site and branch-only tests) have been developed to survey ancient events of adaptive evolution along promoter regions and non-coding elements in apes and other mammals. An important challenge in detection of positive selection is the identification of appropriate and representative neutrally evolving sequences. Differences in mutation rate caused by demographic events may influence the significance of the signatures of adaptive selection for a given DNA element. In other words, faster neutrally-evolving genomic regions may obscure elements evolving by positive selection, while slow evolving neutral regions may overestimate the significance of a weak or neutrally evolving DNA element. The genome assemblies of humans and other great apes have been improved considerably in the past few years. Likewise, new computational methods have been able to align these genomes with more accuracy. A major remaining challenge, however, is identifying sufficient neutral proxy sequences lying outside of known functional elements. Here, we used functional genomics data sets (ATAC-seq) from primate brains and fibroblasts to identify putative regulatory elements. We then mask all known functional elements, drawing on data from the ENCODE and GTEx project. We measure the effect of estimating selection by sampling the remaining putatively neutrally evolving alignments at both the global and local level. More specifically, we compared each open-chromatin site against a random genome-wide alignment of putatively non-functional regions evolving at an average rate, and against at least 20 flanking alignments within 10Kb, 40Kb and 100Kb windows using likelihood approaches. We argue that the combination of these methods can increase the sensitivity by modulating the conservativeness of this test, and allow one to test putative regulatory elements despite being located within regions that are highly masked for exons, introns and other known regulatory elements.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG4

Epigenetic Mechanisms Underlying the Evolution of Learned Vocal Behavior

Morgan Wirthlin [1], Andreas Pfenning [1,*]

[1]Computational Biology Department, Carnegie Mellon University, Pittsburgh, United States

Abstract: Complex behaviors, such as vocal learning (the ability to actively modify vocal production in response to auditory feedback, as in speech), require tight organization of gene activity across a vast array of interconnected cell types and tissues. Tracing the evolution of gene regulation associated with behavior has proven challenging, but by integrating novel computational and experimental comparative approaches, we are beginning to make inroads into the problem. We previously demonstrated that the convergent evolution of learned vocal behavior in humans and birds, which also show convergent specializations in brain circuits devoted to this behavior, also show convergent specializations at the level of gene expression. In the brain regions that control the production of learned vocalizations, we identified convergent changes in gene expression that were unique to vocal learners (humans, songbirds, parrots, and hummingbirds), and not present in their vocal non-learning relatives (non-human primates, doves, and fowl). Given that convergent evolution of behavior is associated with convergent specializations in gene expression, we sought to discover whether this convergence also existed at the level of gene regulation, specifically within the domain of epigenetic modification of enhancer sequences.

To study enhancer regions across large evolutionary distances, we built and validated a specialized pipeline. First, we used multiple sequence alignments to map a set of ~60,000 known the brain enhancers to the genomes of 100 different vertebrate species. In the orthologous enhancers of chimpanzee, macaque, and zebra finch, we estimated enhancer activity using ChIP-Seq for an associated chromatin modification (H3K27ac) in two brain regions, the cortex and the striatum. We found strong conservation of cortical vs. striatal enhancer specificity between humans and non-human primates. Surprisingly, this cortical vs. striatal enhancer activity also showed conservation between human and zebra finch, despite the extensive nucleotide turnover occurring over the >300 million years of evolutionary distance between birds and mammals.

Using this framework, we next identified enhancers found near critical genes with specialized expression in vocal learning birds and humans, that also show specialized enhancer activity in speech motor cortex of humans. Remarkably, investigations of this set revealed that the change in enhancer activity correlated with the specialized transcriptional changes in brain gene expression in vocal learners. To validate and refine our models of enhancer evolution, we have taken several steps to adapt a massively parallel reporter assay for *in vivo* use.

In sum, our work demonstrates that the convergence in behavior, neuroanatomy, and gene expression also extends to convergence in the epigenetic domain, supporting enhancers as a prime candidate for driving the evolution of complex behavior.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-276

Male x Female postcopulatory interactions and the evolution of gametic incompatibility

Yasir Ahmed-Braimah [1,*], Andrew Clark [1]

[1]Molecular Biology and Genetics, Cornell University, Ithaca, United States

Abstract: The genetic basis of postcopulatory sexual selection is poorly understood, even though gametic interactions are essential to reproductive fitness. Importantly, postcopulatory sexual selection processes can drive formation of new species by establishing barriers to fertilization. Here we use the *Drosophila virilis* sub-group as a model to investigate the genetic basis of gametic incompatibilities between species. First, using RNA-seq of adult male reproductive organs, we identify candidate genes that are part of the seminal fluid secretions. Seminal fluid proteins (SFPs) are know to affect a variety of post-mating responses in the female, and are the main paternal effectors in pre-fertilization gametic interactions. We show that these SFPs are rapidly evolving between species, and that the most rapidly evolving SFPs reside within genomic regions that cause paternal incompatibility between species. Second, we examine the regulatory response in female mated to con- and heterospecific males to identify genes that show abnormal postmating responses. We find that heterospecific males transfer testis-specific RNAs, and induce abnormal upregulation of several serine-protease inhibitors in the reproductive tract of females in the first 6 hours after mating. Although we find that the majority of female-specific reproductive genes do not evolve rapidly, one of the misregulated, reproductive tract-specific transcripts evolves rapidly, and resides within a region that contributes to gametic incompatibility in interspecific crosses. We are currently using CRISPR-Cas9 knock-outs to examine the functions of these candidate genes, in addition to using GFP-tagged sperm to understand the functional roles these genes play.

Expanded summary*: Postcopulatory sexual selection is a potent evolutionary force that can drive rapid divergence of reproductive genes. In polyandrous species—where a female can store sperm from multiple males—the opportunity exists for either sperm competition among rival males or preferential fertilization through cryptic female choice. Both of these processes are pervasive in insects and many polyandrous mammals, and are thought to be potential drivers of speciation. While behavioral ecologists have elucidated much of what is known about postcopulatory processes in many taxa, little is known about the molecular genetic mechanisms that control these processes. With the advent of high-throughput sequencing, shot-gun proteomics, and accessible molecular tools for non-model organisms, the genetic basis of postcopulatory sexual selection can now be interrogated with rigor in a wide variety of genetically tractable species.

My research integrates genetic, molecular and bioinformatic approaches to understand the evolutionary dynamics of postcopulatory interactions between males and females. In particular, my goal is to identify important genes that mediate postcopulatory interactions, how these genes are affected by variation within species, the consequences of divergence of these genes between species, and the postcopulatory functions of these genes. To accomplish this, I use the virilis group of *Drosophila* as a model because this species group rapidly evolves interspecific gametic incompatibilities that deem heterospecific sperm incapable of fertilization. Thus, this model system allows detailed investigation of the genetic mechanisms that cause gametic isolation—also known as postmating prezygotic reproductive isolation—and provides an ideal system to characterize the genetic basis of postcopulatory sexual selection processes within species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-269

Exploring the Evolution of Hermaphroditism in Two Species of Gobiid Fish using Whole Transcriptome Sequencing and Gene Coexpression Networks

Jessica Maxfield*, Kathleen Cole [1]

[1]Biology, University of Hawaii Manoa, Honolulu, United States

Abstract: Functional hermaphroditism describes the ability to transition between ova and sperm production at some point in adult life, and has evolved independently numerous times across a broad range of teleost fish taxa. While many aspects of hermaphroditic sexual lability have been explored, we know very little about the gene networks that regulate functional shifts in gamete production. Knowledge of these gene networks would provide a rich source of information from which to propose hypotheses regarding the evolution of functional hermaphroditism in fishes. The family Gobiidae is the ideal system in which to explore both the evolution and development of sexual lability. This taxon is the second largest vertebrate family, and it has been hypothesized that hermaphroditism has arisen independently as many as five times within this group. This study examines shifts in sexual function from ova to sperm production in the marine gobies, *Eviota epiphanes* and *Lythrypnus dalli*, from both a morphological and a molecular perspective. We have performed whole transcriptome sequencing on the gonads of these fishes as they transitioned between gamete types. With these data we have been able to make comparisons between these species with regard to what genes are being used to regulate these shifts, when they are being upregulated, and the levels of gene expression. These data will be used to construct gene coexpression networks to provide insight into the gene regulatory pathways that govern changes in sexual function and to test the hypothesis of the independent origin of hermaphroditism with respect to these two lineages.

Expanded summary*: Mechanisms responsible for the development of specific sexual phenotypes (i.e. male, female, hermaphrodite) show remarkable diversity throughout the animal kingdom. Variation in factors important in the development of sexual phenotypes can be a major driver of speciation. For example, in many animal taxa strong positive selection has been found for numerous sex-related genes, which also undergo accelerated rates of evolution. Despite the central importance of sexual phenotypes in animals, studies of the underlying mechanisms associated with sexual development have been confined to a small number of model organisms, which are not sexually plastic. This offers a limited view of the evolution of sexuality and fails to elucidate broader patterns and innovations in sexual development.

Teleost fishes show the greatest diversity and flexibility in sexual systems and sexual expression among vertebrates. Functional hermaphroditism in which individuals produce both ova and sperm at some time during adult life, in particular, has been the focus of a large body of research. The mechanisms underlying the ability to transition from producing one gamete type to another, including the gene networks that regulate this transition however, are not well understood.

My dissertation research goals are to develop a genetic model for sex reallocation using next generation sequencing in the hermaphroditic gobiid species, *Lythrypnus dalli* and *Eviota epiphanes.* These species represent two gobiid lineages which are proposed to have independently-derived hermaphroditism, based on their phylogenetic relationships and their unique and differing gonad morphologies. These two species are found in different clades, sharing their most recent common ancestor over 40 million years ago, and there are many intervening lineages of non-hermaphroditic species between them.

The process of sex reallocation will be characterized in these two goby species by comparing concurrent changes in gonad morphology and gene expression as individuals transition between ova and sperm production. Co-expression networks will be generated from RNAseq data for each species and compared, to determine to what extend they exhibit differences in gene function and gene interaction associated with gamete production. A finding of identical regulatory processes between these two species will support a conservation of evolved mechanisms associated with shifts in gamete production, while a finding of differing regulatory processes will support a hypothesis of divergence in underlying regulatory mechanisms

This study will establish a methodology for future studies on sexual development of both fish and other vertebrates. The overall focus of this study is the use of a comparative approach to address questions surrounding the evolution of novelty and speciation, through the examination of reproductive processes associated with repeated shifts in sexual function. Hypotheses regarding the evolution of hermaphroditism have been proposed for fishes, but have never been tested at the gene regulatory level. This study will be testing one

such hypothesis, using molecular methods, in the family Gobiidae, and will establish a methodology for similar hypothesis-testing approaches in other taxa.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-279

piRNA diversity and abundance is dependent on transposable element dynamics in mammals

Mike Vandewege [1], Neal Platt [1], Aliceanne Szeliga [2], David Ray, Federico Hoffmann [3,*]

[1]Dept. of Biological Sciences, Texas Tech University, Lubbock, [2]Dept. of Biology, Harvey Mudd College, Claremont,

[3]Dept. of Biochemistry and Molecular Biology , Mississippi State University, Mississippi State, United States

Abstract: PIWI proteins and PIWI interacting RNAs (piRNAs) are part of a cellular pathway that protect genomes against the proliferation of transposable elements (TEs). PIWIs and piRNAs assemble into complexes that are involved in epigenetic and post-transcriptional repression of TEs, where piRNAs identify targets for these complexes via sequence complementarity. TE silencing occurs via a feed-forward amplification loop known as the “ping-pong” cycle. Here, we compare piRNA repertoires, PIWI expression, and TE expression throughout development in rabbit, mouse, and ground squirrel, a mammal with very low levels of TE activity to gain insights into the interplay between piRNAs, PIWIs and TEs. Specifically, measured the expression of PIWI proteins and their piRNA counterparts to ask how piRNA responses relate to variable TE content. Our results suggest major differences in the temporal expression of PIWI in the ground squirrel compared to rabbit and mouse. We also identified temporal variation of TE expression in the rabbit and mouse; temporal changes were absent the ground squirrel. Lastly, we found an increase in the ping-pong response resulted in a decrease of TE transcription in the mouse and rabbit, but this observation was absent in the squirrel. Our major findings indicate that in a genome with a limited TE threat, PIWIs incorporate sense and anti-sense TE transcripts into a ping-pong pathway but do not appear to reduce TE transcription.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-267

Modeling the evolution of coding sequence A-to-I RNA-editing

Mengyi Sun [1,*], Jianzhi Zhang [1]

[1]Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, United States

Abstract: A-to-I editing refers to the cellular process that enzymatically converts adenosines to inosines in RNA molecules. Because ribosomes interpret inosines as guanines, RNA editing in coding regions could lead to amino acid changes in the encoded proteins. Previous transcriptome studies revealed diverse patterns of A-to-I editing across species. In humans, the frequency of nonsynonymous editing relative to that of synonymous editing (F_N/F_S) is lower than 1, and the median level of nonsynonymous editing relative to that of synonymous editing (L_N/L_S) is also lower than 1. In squid and fruit fly, $F_N/F_S < 1$ but $L_N/L_S > 1$. In *Fusarium* fungi, F_N/F_S and L_N/L_S both exceed 1. To understand why F_N/F_S and L_N/L_S vary substantially among species, we modeled the evolution of RNA editing by considering the editing level of a site as a quantitative trait. Our model includes parameters of mutation rate, mutation size, natural selection, and population size, and can more or less recapitulate the observe patterns of RNA editing in different species. We found that positive selection for RNA editing has to be invoked to explain $F_N/F_S > 1$; the same is true for $L_N/L_S > 1$. Nevertheless, we also found that unbiasedly estimating F_N/F_S and L_N/L_S can be difficult because of potential mutational biases in creating editing substrates. These results cation evolutionary interpretations of F_N/F_S and L_N/L_S estimates, but support the role of positive selection in shaping RNA editing in multiple species provided that the current estimates are unbiased.

Expanded summary*: A-to-I editing refers to a cellular process that converts adenosine to inosine in RNA molecules. Because translational mechanisms interpret the inosine as guanine, RNA-editing in the coding region could alter the codon and lead to amino acid changes in the encoded protein. This feature enables organisms to use different isoforms of proteins in different tissues and different developmental stages, which adds a dimension to gene regulation. Moreover, RNA-editing in a particular site is generally not 100%, and one gene could harbor multiple RNA-editing sites. Thus, RNA-editing potentially could produce a large number of isoforms from the same gene. Together with alternative splicing, RNA-editing might contribute to proteome diversity and organismal complexity. Given that RNA-editing has taken part in many important cellular processes, it is not surprising that mis-regulation of RNA-editing has been reported to cause diseases. For instance, loss of RNA-editing in a glutamate-gated channel gene leads to abnormal behavior in mouse. Recent work on cancer cell lines and fly development implies RNA-editing may play important roles in carcinogenesis and developmental processes.

The fast development of high-throughput RNA-sequencing now allows us to profile RNA-editing genome widely. Previous studies of genome-wide RNA-editing in several species reported drastically different patterns measured by F_N/F_S (the frequency of nonsynonymous editing relative to that of synonymous editing) and L_N/L_S (the median level of nonsynonymous editing relative to that of synonymous editing), respectively. To understand the exact evolutionary interpretation of these measures, we modeled the evolution of RNA editing by considering the editing level of a site as a quantitative trait. Our model recapitulated the observed patterns qualitatively. More over, we found that positive selection is necessary for either measure to be larger than 1, provided that the estimation of F_N/F_S and L_N/L_S is unbiased. We also dicuss the conditons that might lead to biased estimation of F_N/F_S and L_N/L_S.

Our result might also provide some insights to phenotypic evolution. For instance, we found that the variation of effective population size (*Ne*) does not contribute to the variation of F_N/F_S and L_N/L_S, which is unexpected based on the drift-barrier hypothesis. This result indicates that the generalization of the conclusions from molecular evolution to phenotypic evolution should be more cautious.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG11

Integrated analysis of open chromatin and gene expression data identifies gene regulation changes in decidual cell evolution

Cong Liang [1 2,*], Arun Chavan [1 3], Eric Erkenbrack [1 3], Gunter Wagner [1 3]

[1]Systems Biology Institute, Yale University, West Haven, [2]Program of computational biology and bioinformatics, [3]Department of Ecology and Evolutionary biology, Yale University, New Haven, United States

Abstract: Evolution of gene regulation plays an important role in the origin of novel cell types. However, gene regulation discoveries are mostly built upon single gene perturbations, which are difficult to implement on a genome-wide scale. The development of the assay for transposase-accessible chromatin using sequencing (ATAC-seq) provides an opportunity to infer open chromatin with a small amount of starting material in a single experiment. Open chromatin regions identified by ATAC-seq, usually hundreds of base pairs, are available for transcription factor (TF) binding. But the sequencing signal suffers from low abundance at single base pair resolution to accurately predict specific TF binding footprints. We introduce an integrated method utilizing open chromatin data from ATAC-seq, gene expression data from RNA-seq, as well as the known binding site information, to identify the gene regulation change between two cell states. We apply the method to human decidualized stromal cells that differentiate from endometrial stromal fibroblasts to identify the induced regulations during decidualization. We compared our findings in human endometrial stromal cells with a homologous cell type in the opossum, which are thought to lack the decidualization mechanism. We discovered that some induced regulation already exist in opossum, and identified human specific decidualization regulations. We confirmed these findings with perturbation experiments. This research integrates multiple data resources to infer gene regulatory change between two cell states, also proposed and tested a hypothesis for how these states have evolved. Our results present important mechanisms for understanding cell type evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-278

Meta-analysis reveals details about the environmental stress response in budding yeast.

Keffy Kehrli [1,*], Joshua Rest [1]

[1]Ecology and Evolution, Stony Brook University, Stony Brook, United States

Abstract: Changes in gene expression patterns across environments are essential to regulate cellular processes, especially in single-celled organisms that are in direct contact with environmental conditions. The yeast environmental stress response (ESR) is a set of characteristic gene expression changes after exposure to various environmental stressors, such as heat shock, oxidative stress, and osmotic stress.

Existing microarray data has been frequently carried out with few biological replicates, limiting the statistical power of these results. Here, we use meta-analysis and meta-regression to combine many transcriptomic datasets in statistically robust ways to find consistent patterns across studies, account for variables such as study design and batch effects, and weight studies according to their power and variance. We used meta-analysis to evaluate the consistency of ESR genes across many studies, evaluate whether there are sets of ESR genes that are characteristic of classes of stressors, and to discover low expressed genes in the ESR that would only be evident by greatly increasing statistical power.

The meta-analysis was conducted on gene expression data from a large collection of yeast gene expression microarray experiments that measure responses to various stressors. We detected statistically significant differences in gene expression between treatment conditions in low expression genes that were not identified in any individual study. We also determined transcriptional responses that are specific to differences in how stresses were applied, resulting in a more tightly defined general ESR. Finally, we identified differences in the ESR that can be specifically attributed to genetic differences among strains.

Expanded summary*: The yeast environmental stress response (ESR) is a common response to various environmental stressors, such as heat shock, oxidative stress and osmotic stress. Stress responses are key to organisms adjusting to and surviving changing environmental conditions. In yeast, the environmental stress response is a generalized response that is entered into via responses to different stresses. *S. cerevisiae* that has been exposed to one stress and survive are more resistant to large doses of other stresses via the ESR. Further research into this regulatory process provides insight into the mechanisms by which organisms adapt to changing environments and survive environmental processes that they have not previously experienced.

Micro-array data has been frequently carried out with few biological replicates, limiting the statistical power of these results. It is therefore difficult to determine exactly how much of the data is accurate, especially when individual experiments have conflicting results. Experiments to probe gene expression changes during stress responses in yeast have implicated nearly two-thirds of the genome in individual experiments, but not all genes in all experiments. Meta-analysis and meta-regression methods provide a statistically robust method to determine which genes have differential gene expression during stress responses in yeast.

Meta-analysis and meta-regression is also broadly applicable to microarray data from other organisms. Previous work applying meta-analysis to gene expression data has contributed to understanding of water stress in *Arabidopsis,* and can be applied to microarray data from a number of other organisms, including wild isolates of *S. cerevisiae* and related species of yeast. This can provide insight into the evolution of stress responses, including the ESR, which has been found in *S. cerevisiae* and *S. pombe* (first described in *S. cerevisiae*), but not in *C. albicans.*

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POB-422

ELFA, A NEW PLATFORM FOR THE DIGITAL EVOLUTION OF GENE REGULATORY NETWORKS

Anselmo Pontes*

Poster: Evolutionary computation recreates Darwinian evolution in computer algorithms and is an invaluable tool in experimental evolution. For decades, it has helped uncover selective pressures and evolutionary processes that would otherwise be difficult or impossible to observe in a wide range of fields, including ecology, animal behavior, microbiology and neuroscience. However, the questions we can address using evolutionary computation are constrained by the choice of software and its genetic representation, which often limits the investigation to a high level of abstraction. One subject that has been particularly challenging is the evolution of the gene regulatory networks responsible for cell decision-making. We will introduce a new digital evolution platform built upon a mechanistic model of gene regulation and protein signaling, that permits studying the evolution of an organism's adaptive decision-making and its underlying gene regulatory and signaling networks. We will discuss how this platform is being used to investigate the evolution of homeostasis and circadian rhythms in prokaryotes, and how we can apply it to eukaryotic organisms and address questions related to the evolution of multicellularity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-274

Investigating the potential roles of long non-coding RNAs as splicing modulators through RNA:RNA interactions

Elżbieta Wanowska [1,*], Michał Szcześniak [1], Magdalena Kubiak [1], Izabela Makałowska [1]

[1]Department of Integrative Genomics, Institute of Anthropology, Faculty of Biology Adam Mickiewicz University , Poznań, Poland

Abstract: Long non coding RNAs (lncRNAs) constitute a large group of transcripts that are over 200 nucleotides, without protein coding capacity. They have been shown to play important roles in various biological processes and, in particular, they function as regulators of gene expression, both during the act of transcription and post-transcriptionally. They are also implicated in a number of human diseases, especially cancers. However, the functions of most lncRNAs remain unclear. Still little is known about their engagement in gene expression regulation through lncRNA:RNA interactions. By hybridizing with other transcripts, lncRNAs could be involved in at least several regulatory mechanisms, including modulation of splicing, triggering RNA editing events, guiding protein-coding transcripts to degradation in a Staufen-mediated decay pathway, and abrogation of miRNA-induced repression by masking miRNA target sites.

Our goal is to investigate splicing-associated functions of lncRNAs that are exerted in the context of RNA:RNA duplexes. Recently, based on *in silico* predictions of lncRNA:RNA base-pairings across the human transcriptome, we discovered that there is a great potential for lncRNAs to play a role of splicing modulators. This could be achieved by masking splice sites and other splicing signals. We are now focusing on possible roles of lncRNAs in splicing modulation that results in a shift in coding capacity transcripts. We have selected our candidates according to transcripts biotype, protein coding capacity of genes and lncRNA:RNA interaction regions. We have chosen lncRNAs detected at higher level in the nucleus than in the cytoplasm, based on bioinformatics analyses. At the moment, we are experimentally testing the most promising candidates. The first experiments involve cell fractionation and Real-Time PCR in HEK293 cell lines to confirm our bioinformatics predictions. Then, silencing of nuclear lncRNAs will be performed and effect on splicing of potentially regulated genes will be determined. Our study should add up to better understanding of lncRNA biology and help decipher their functions in the context of RNA:RNA interactions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-283

Fitness contributions and in vivo function of the tandem glycine riboswitch in Bacillus subtilis

Arianne Babina [1,*], Nicholas Lea [1], Michelle Meyer [1]

[1]Biology, Boston College, Chestnut Hill, United States

Abstract: In many bacterial species, the glycine riboswitch is comprised of two homologous RNA ligand-binding domains (aptamers) that each bind glycine and act together to regulate expression of genes involved in glycine metabolism. While the structure and molecular dynamics of the tandem glycine riboswitch have been the subject of numerous *in vitro* studies, the purpose and selective advantages of the dual aptamer architecture and the *in vivo* behavior of the riboswitch remain unclear. To examine the proposed models of tandem glycine riboswitch function in a biologically relevant context, we characterized the regulatory activity of mutations to the riboswitch structure using beta-galactosidase reporter assays in *Bacillus subtilis*. To assess the impact disruptions to riboswitch function have on cell fitness, we introduced these mutations into the native locus of the tandem glycine riboswitch within the *B. subtilis* genome and assayed the mutant strains under a variety of conditions. We find that mutations to the leader-linker kink-turn and glycine-binding and dimerization domains disrupt riboswitch regulation, reduce gene expression, and inhibit swarming motility and biofilm formation in high glycine environments. Our results suggest that all of the previously described interactions contribute to tandem glycine riboswitch regulation *in vivo* and that both aptamers are required for proper riboswitch function and maximal gene expression. In *B. subtilis*, expression of the glycine riboswitch-regulated *gcvT* operon is necessary for complex cell behaviors in high glycine conditions. Our experimental approach offers a novel way to explore the physiological roles of riboswitches within the context of their native loci.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-268

Expression quantitative trait loci of dosage-sensitive genes have narrow tissue specificity bias

Alan M. Rice [1,*], Pauric Donnelly [1], Aoife McLysaght [1]

[1]Trinity College Dublin, Dublin, Ireland

Abstract: Dosage-sensitive genes are often seen to be refractory to variation. Ohnologs, paralogs produced from whole genome duplication events at the base of the vertebrate lineage, and haploinsufficient genes have been shown to be depleted on human benign copy number variants but enriched on pathogenic variants. This intolerance to copy number change is likely due to an expression constraint encountered by some variants. Expression quantitative trait loci (eQTLs) are genomic regions harbouring sequence variants that influence the expression level of one or more genes within various tissues. Contrary to our expectation that dosage-sensitive genes should be depleted for eQTLs, we find that they are in fact enriched for this variation and that these eQTLs are biased towards having narrow tissue specificity. Dosage-sensitive genes have fewer eQTL-affected tissues than dosage-insensitive genes, as their eQTLs are more tissue-specific with broad tissue breadth eQTLs likely removed by purifying selection due to conflicts with expression constraints. Additionally, we observed that ohnolog pairs have more similar eQTL-affected tissues compared to random ohnolog pairs suggesting a shared constraint between real pairs likely imposed by dosage-balance. These patterns suggest that dosage-sensitivity shapes the evolution of eQTLs influencing the expression of constrained genes whereby deleterious variants in conflict with constraints experience purifying selection.

Expanded summary*: Expression quantitative trait loci (eQTLs) are genomic regions harbouring sequence variants that influence the expression level of one or more genes [1]. A range of eQTL effect sizes, both increasing and decreasing expression, are observed and can occur across a number of tissues or, more typically, in a tissue-specific manner [2]. In human, the expression of thousands of genes is affected by eQTLs making them a significant contribution to the genetic variation of expression and in turn phenotypic variation and complex disease. The majority of the genome, therefore, must be able to tolerate some amount of mRNA level change without apparent deleterious consequences. However, in combination with genome-wide association studies, eQTLs have been used to elucidate further the pathophysiology of many disease phenotypes. To date, eQTLs have been associated with human diseases including asthma, autoimmune disorders, diabetes, numerous cancers, Parkinson's disease, and other brain disorders (see Table 1 in ref [1]). Therefore, the effect of eQTLs on gene expression and association with important traits makes them worthy of study especially in the context of genes with known expression constraints.

Dosage-sensitive genes are often seen to be refractory to variation. In human, ohnologs, paralogs retained after whole genome duplication events at the base of the vertebrate lineage have been shown to be depleted on control and benign copy number variants (CNVs) but enriched among genes on pathogenic variants [3, 4, 5]. Likewise, similar trends are observed for genes without gene duplication or loss events in mammalian genomes and also haploinsufficient genes [6]. This intolerance to copy number change is likely due in part to a constraint on expression that is encountered by some variants. When a variant arises such as a CNV or eQTL that causes a deleterious aberration in expression level, the variant will experience purifying selection and be removed from the population. Therefore we expect that dosage-sensitive genes are affected less by eQTLs and that the human genome and its segregating variants should contain the hallmarks of this selection.

Here, we investigated the patterns of eQTLs affecting dosage-sensitive genes. Contrary to our expectation that ohnologs and other categories of dosage-sensitive genes should be depleted for this variation, we found that these genes are enriched for eQTLs. However, they have fewer eQTL-affected tissues than dosage-insensitive genes, as the eQTLs that affect these genes are more tissue-specific. Furthermore, controlling for gene age, we still find that ohnologs have more constrained eQTL patterns compared to nonohnologs of the same age. Dosage-sensitive genes are depleted for broad-tissue breadth eQTLs, likely because broad-tissue breadth eQTLs will conflict with constraints more often, giving rise to deleterious expression levels. We found that eQTLs that affect multiple genes have graduated constraint increasing with larger proportions of dosage-sensitive genes affected. While most eQTLs affect a single gene, it is noteworthy that the inclusion of dosage-sensitive genes as part the group of genes influenced by multi-gene eQTLs has consequences on the variation affecting dosage-insensitive genes. However, ohnologs are disproportionately more constrained than nonohnologs when affected by the same multi- gene eQTL. We observed that ohnolog pairs have more similar eQTL-affected tissues compared to random ohnolog pairs suggesting a shared constraint between real pairs.

This evidence suggests that dosage-sensitivity shapes the evolution of eQTLs influencing the expression of these genes whereby deleterious variants in conflict with constraints experience purifying selection. Patterns of selection acting on eQTLs are likely due to functional constraints and have important implications for the identification of candidate disease-causing variants that affect dosage-sensitive genes.

[1] Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. 16, 197–212 (2015).
[2] Gerrits, A. et al. Expression quantitative trait loci are highly sensitive to cellular differentiation state. PLoS Genet. 5, e1000692 (2009).
[3] Makino, T. & McLysaght, A. Ohnologs in the human genome are dosage balanced and frequently associated with disease. Proc. Natl. Acad. Sci. U.S.A. 107, 9270–9274 (2010).
[4] Makino, T., McLysaght, A. & Kawata, M. Genome-wide deserts for copy number variation in vertebrates. Nature Communications 4, 2283 (2013).
[5] McLysaght, A. et al. Ohnologs are overrepresented in pathogenic copy number mutations. Proc. Natl. Acad. Sci. U.S.A. 111, 361–366 (2014).
[6] Rice, A. M. & McLysaght, A. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. Nature Communications 8, 14366 (2017).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG10

Predicting specific biological functions of gene regulatory regions using convergent evolution

Raghavendran Partha*, Maria Chikina, Nathan Clark

Abstract: A recent explosion of genome-scale efforts has identified a wealth of active regulatory elements in the mammalian genome. However, assigning higher level functional roles for these elements remains a major challenge. Computational tools learning patterns of convergent evolution based on comparative genomics data offer a unique opportunity to generate accurate hypotheses for the specific roles of these elements in gene regulation. Previous and ongoing work has proven the strength of these approaches in identifying convergent changes in protein-coding genes as a result of adaptation to similar environmental pressures.
Here, we present a novel evolutionary rates-based approach that detects convergent changes at non-coding regions associated with phenotypic adaptations. As an initial application, we study the convergent adaptation of four mammals to the subterranean environment. Based on the accelerated rates of evolution on the subterranean branches, we demonstrate that our method successfully identifies enhancers controlling the expression of the transcription factor Pax6 in eye amidst non-eye enhancers and other uncharacterized non-genic regions.
We performed benchmarking analysis using simulations to identify the most powerful variant of our method that consistently detects convergent shifts in rates. Simulation conditions were varied across several parameters, including the number of convergent adaptations.
We finally validate our best performing method on large-scale datasets including i. predicting eye-specific enhancers in the VISTA enhancer database and the ENCODE project; ii. prioritizing candidate regulatory elements near eye development genes. Our method detects a strong enrichment of convergence near annotated eye-specific elements, and further reveals multiple uncharacterized elements showing strong convergence. These novel elements are thus candidates for involvement in transcriptional regulation during ocular development, and can help focus experimental efforts aimed at identifying genetic mutations underlying eye-related disorders.
Overall, our analyses showcase the potential of convergent evolution-based tools to provide functional annotation for regulatory regions in the mammalian genome.

Expanded summary*:

Significance

Phenotypic and genotypic variation between species are the result of millions of experiments performed by nature. Technological advancements over the past two decades have paved the way for rapid generation and assimilation of multi-dimensional data that hold the key to a more comprehensive understanding of why and how this variation arose. On the other hand, the availability of this new wealth of data poses a much more challenging problem - developing tools that can successfully unlock the association between variation at the phenotypic level to that at the level of the genotype. To develop effective solutions for a problem of this scale, there is a need for inter-disciplinary efforts that combine insights from complementary fields of research.

The survival of a species is contingent on its successful adaptation to the specific challenges in its environment. The rich diversity of species in nature has provided countless examples of multiple unrelated lineages showing phenotypic adaptation to similar environmental challenges. This opens up the opportunity to develop evolutionary-based approaches aimed at inferring the changes at the genetic level underlying said instances of phenotypic convergence. Alternatively, genetic elements showing convergent changes in species characterized by phenotypic convergence are thus strong candidates for a functional role in the phenotype. Successful applications of these approaches on whole genomes can therefore generate accurate hypotheses for the specific functional roles of large numbers of genetic elements.

Results

The independent transitions to a subterranean lifestyle of four mammals present one such example of convergent phenotypic adaptation. Application of our evolutionary rates-based approach on a set of mammalian genomes reveals large numbers of genes convergently accelerated in these subterranean mammals. Genes showing strong rate acceleration in these mammals are highly enriched for function in visual perception. This effect is the result of a relaxation of constraint on the visual pathways in these mammals in their dim-light underground environment, thus leading to regression of eye-specific genetic elements. We also developed

a novel variant of our rates-based method to test if non-coding regions show convergent regression in the subterranean mammals. By applying this new method on conserved non-coding regions near *Pax6*, we demonstrate that eye-specific enhancers of *Pax6* show convergent rate acceleration in the subterranean mammals in comparison to non-coding regions with no eye-specific annotation. The example illustration of the power of our method in detecting eye-specific enhancers of *Pax6* shows the promise of extending our approach to perform genome-wide scans for candidate non-coding elements involved in ocular development. We expect the results of our scans to reveal prime candidates of interest for biomedical researchers experimentally investigating genetic elements associated with eye-related developmental disorders.

Future work

We strive to make our methods more broadly applicable, outside of just vision-related elements. In order to test the power of our method to identify convergent rate shifts underlying phenotypic adaptation across a range of evolutionary scenarios, we perform benchmarking analyses using simulations. We identify the most optimally performing variant of our method across conditions with varying parameters, including the number of convergent adaptations. Examples of other traits that we are currently investigating include lifespan in mammals, hindlimb presence/absence etc.

Impact

Ultimately, we seek to distribute our tools as publicly available software, useful for the general research community to perform genome-wide scans for candidate regions underlying specific phenotypic adaptations. Researchers will be able to investigate their trait of interest and identify genomic regions showing correlated rates of evolution with their trait based on genomic sequences of 62 mammalian species. Following future developments, the tools will further become applicable to non-mammalian evolutionary lineages such as fungi, nematodes etc. As more genomes are sequenced, we anticipate a concordant increase in the interest generated by, and the power of computational efforts learning from patterns of evolution to reveal gene regulatory changes underlying convergent phenotypes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-272

Evolutionary dynamics of alternative splicing regulation in mammalian development

Pavel Mazin [1,*], Margarida Cardoso-Moreira [2], Philipp Khaitovich [1], Henrik Kaessmann [2]

[1]Center for Data-Intensive Biomedicine, Skolkovo Institute of Science and Technology, Moscow, Russian Federation,

[2]Center for Molecular Biology, Heidelberg University, Heidelberg, Germany

Abstract: Alternative splicing (AS) is one of the major mechanisms of proteome diversification in eukaryotes, especially in mammals. Recent studies have shown its crucial involvement in normal development, stress response, and disease. Contrary to the evident functionality of AS, interspecies comparisons revealed an overall low conservation of AS in adult organs. To solve this puzzle, we investigated the evolution of AS regulation during embryonic and postnatal development of seven tissues in six mammals and a bird (red jungle fowl) based on RNA sequencing data for a total of approximately 2,000 samples. Our results show that developmental AS changes are highly tissue-specific and mostly occur in brain, heart, and adult testis. Interestingly, we discovered two distinct brain-specific regulatory programs that are specific to pre- and post-natal stages, respectively. In most cases, we observed two tissue-specific patterns: the first one corresponds to tissue-specific exon inclusion, the second one to tissue-specific exclusion. Strikingly, inclusion patterns are associated with substantially more regulatory elements than exclusion patterns. In brain and heart these mirrored patterns also exhibit mirrored positioning (relatively to the exon) of regulatory elements in pre-mRNAs. We discovered an overall strong evolutionary conservation of these developmental splicing programs across species. The aforementioned patterns exhibited conserved association with many specific sequence motifs. Most of them were not associated with any known RNA-binding proteins. Analysis of species-specific AS revealed two sources for alternative exons: the “alternification” of previously constitutive exons, which is most pronounced in brain, and the birth of new exons, which are more abundant in testis. We speculate that the mirrored patterns described above could arise through combinations of these two mechanisms.

Expanded summary*: Alternative splicing (AS) allow single eukaryotic gene to produce multiple mRNA and, consequently, several proteins. AS is very abundant in high eukaryotes, it is believed that 99% of human multi-exon genes undergo AS. Numerous studies have shown its crucial involvement in normal development, stress response, and disease. Contrary to known functionality of AS recent studies have shown rapid evolution of AS with inter-species differences exceeding inter-tissue ones. Studies focused on specific genes have shown the role of AS in cell differentiation but most of genome-wide studies of AS were focused on adult tissues or post-natal development. Thus, the role of AS in embryonic development is mostly unknown. To address these issues we used about two thousands polyA RNA-Seq samples obtained from seven tissues (cerebral cortex, cerebellum, heart, liver, testis, ovary, and kidney) dissected from seven species (human, macaque, mouse, rat, rabbit, opossum, and chicken) of different ages ranging from early stages of embryonic development to sexual maturation. Our data allowed us to investigate tissue-specific differentiation of AS from very early stages when tissues are mostly indistinguishable in most of species. We found multiple patterns of developmental AS changes. The most abundant patterns were specific to brain, heart, or adult testis. Strikingly, almost the same sets of patterns were discovered in all species independently. Comparison of similar patterns across species revealed that in most cases they encompass orthologous exons. Thus, we discovered strong conservation of AS developmental regulatory programs. Most of tissue-specifc patterns were found in two versions: the first one corresponds to tissue-specific exon inclusion, the second one to tissue-specific exclusion. These mirrored patterns are differ in many properties: inclusion patterns are usually more conserved, more abundant and have more regulatory sequences in adjacent introns. Additionally, in brain and heart these mirrored patterns also exhibit mirrored positioning (relatively to the exon) of regulatory elements in pre-mRNAs. We compared different species and discovered two major directions of AS evolution: (i) “alternification” of constitutive exons which is most pronounced in brain, and (ii) the birth of new exons, which are more abundant in testis. We speculate that the former mostly contribute to the exclusion patterns while the second produce exons for the inclusion patterns. The differences in ancestral sequence background can partially explain the differences observed between these two types of AS patterns. Taking together we prove the conservation of developmental AS patterns and suggest evolutionary mechanism responsible for appearance of the two types of the AS patterns.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-266

Evolution of regulators or targets?

- Evolution of twin microRNAs sheds light on miRNA regulatory roles

Yixin Zhao [1,*], Pei Lin [1], Zhongqi Liufu [1], Hao Yang [1], Xu Shen [1], Chung-I Wu [1 2], Tian Tang [1]

[1]School of Life Sciences, Sun Yat-sen University, GuangZhou, China, [2]Department of Ecology and Evolution, University of Chicago, Chicago, United States

Abstract:

An enduring mystery in microRNA (miRNA) biology in animals is why and how the large number of targets, most of which are only weakly repressed, is maintained. One suggestion is that this situation is transient, and that miRNAs evolve toward fewer but more strongly repressed targets. To test this hypothesis, we studied miRNAs that produce two mature miRs from the same hairpin precursor. These twin-miRs usually regulate non-overlapping sets of targets. Therefore, selection driving a reduction in target pool size, if it exists, should reduce these loci to solo-miRs making twin-miRs evolutionarily transient and uncommon. We conducted a systematic survey of 45 small RNA libraries from several tissues of *Drosophila melanogaster*. We filtered the data sets to identify high-confidence twin-miRs and studied their evolutionary persistence and regulation. To our surprise, we found that one third of the *D. melanogaster* miRNAs are twin-miRs, with both precursor arms nearly equally expressed in some tissues. These twin-miRNAs are common in both phylogenetically old and young sets of miRNAs and are functional, conserved, and exhibit expression patterns comparable to solo-miRs. Our results suggest that miRNAs may not be driven by selection toward a smaller target pool. We propose that miRNA-mediated maintenance of transcriptional homeostasis leads to evolutionary retention of a large number of targets.

Expanded summary*:

MicroRNAs are the most abundant regulatory molecules on the per gene basis in metazoan cells. Since the top 50 miRNA genes contribute nearly as many transcripts as all mRNA genes combined, they are arguably the dominant regulatory molecules in stoichiometric terms.

The central conundrum about the regulatory role of miRNAs in metazoans is their diffuse actions - each miRNA represses more than 100 target genes, but only weakly. A common hypothesis is that the regulator-targets relationship would coevolve to become stronger. Presumably, the lesser repressed targets would drop off and the remaining targets would become more strongly repressed. However, this prediction is not supported by observations – target number does not decrease with the age of miRNAs (Shomron, et al. 2009; Nozawa, et al. 2016). A possible explanation is that the selective advantage of dropping one target gene out of several hundreds may be too small.

Here, we test the coevolution hypothesis by examining the regulators themselves. Some miRNAs are expressed as twin-miRs with a major and minor form derived from the same hairpin precursor. If the coevolution hypothesis is correct, selection should have driven the evolution toward solo-miRs, which produce only a single mature miRNA from the precursor. By extensively analyzing the miRNA expression data in Drosophila, we find the co-evolution hypothesis to be untenable.

In this study, we experimentally demonstrated that both mature products derived from twin-miRs act to repress target transcripts in *D. melanogaster*. We also analyzed expression and evolution patterns of these miRs, contrasting them to the well-characterized solo-miRs. If the coevolution hypothesis is correct, we would expect that twin miRNAs represent an evolutionary stage when new miRNAs optimize their expression by choosing one arm as the dominant product. Such transitions would be more consequential than single-target elimination and thus should leave clear evolutionary foot-prints. Our analyses of miRNA arm usage in *D. melanogaster* do not support the notion that twin miRNAs are evolutionarily transient and uncommon. Indeed, we find that one third of miRNA genes in Drosophila use both arms to produce mature miRNAs, often in the same tissue types, and this pattern persists through millions of years of evolution. Moreover, minor miRs are often evolutionarily constrained and, when moderately expressed, can be shown to repress target transcripts. It thus appears that the production of twin-miRs is a normal aspect of miRNA biology.

The conclusion is that the diffuse repression by miRNAs is not evolutionarily transient. It is in fact evolutionarily conserved. The result hence lends strong support to the alternative view that miRNAs stabilize gene regulatory networks by broadly but weakly repressing large number of target genes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-275

First depleted, then enriched: the evolution of transposable element regulatory function

Corinne Simonti [1], Mihaela Pavlicev [2], John A. Capra [3,*]

[1]Vanderbilt Genetics Institute, Vanderbilt University, Nashville, [2]Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, [3]Biological Sciences, Vanderbilt University, Nashville, United States

Abstract: Transposable elements (TEs) make up more than half of the human genome. Since TEs contain regulatory sequences that promote their transcription and amplification, they provide a fertile landscape of potential gene regulatory elements on which evolution can act. Indeed, many classes of TE contribute to the evolution of gene regulation in different lineages through co-option into alternative gene promoters, enhancers, and even insulators. However, the current understanding of TE-based rewiring of gene regulatory programs comes from a small number of examples, largely in the immune and reproductive systems.

We comprehensively evaluated the prevalence and evolutionary dynamics of the co-option of TEs into gene regulatory enhancer and promoter elements across 112 cell lines and primary tissues from the FANTOM consortium. Overall, TEs are significantly depleted of regulatory enhancer activity compared to the genomic background ($P < 0.0001$). The degree of depletion varied across contexts (1.5–3x), but it was significant in every cellular context considered. Promoters were even more significantly depleted of TEs than enhancers (2.9x vs. 2.3x overall). Thus, in spite of their regulatory potential, TEs are significantly less active than non-TE regions genome-wide. This suggests that cells actively repress the activity of TE sequences, perhaps to protect themselves from the mutagenic properties of active TEs. Nonetheless, we find that enhancers with tissue-specific activity are significantly more likely to be derived from TEs. Furthermore, the likelihood that a TE has enhancer activity increases with its age. Ancient TEs (originating before the divergence of amniotes) are 9.2 times more likely to have enhancer activity than TEs that integrated on the great ape lineage, and young TE-derived enhancers are significantly more likely to be tissue-specific in activity. Nonetheless, we identified a small number of TE families, most notably the endogenous retroviruses (ERVs), with different dynamics that highlight unique functional trajectories and evolutionary innovations.

Our data suggest striking similarity in the evolutionary and functional dynamics of different TE families. TEs appear to be actively repressed upon integration into the genome, leading to the degradation of their sequences over time. However, when a specific element gains regulatory function in a tissue, it becomes protected from degradation, and may gain activity in other tissues.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG8

The gene expression consequences of mammalian regulatory evolution

Diego Villar [1,*], Camille Berthelot [2], Duncan Odom [1], Paul Flicek [2]

[1]Cambridge Institute - Cancer Research UK and University of Cambridge, [2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, United Kingdom

Abstract:

How stable gene expression is maintained by rapidly evolving collections of enhancers and promoters is a fundamental question in evolutionary genetics.

Here, we evaluate the consequences of regulatory evolution on mammalian gene expression by jointly analysing the activity of promoters and enhancers with downstream transcript levels, measured across the same liver samples from over twenty mammalian species. We confirm largely conserved gene expression levels, and enhanced expression stability for subsets of genes relevant to tissue physiology. Genes associated with complex regulatory landscapes across species generally exhibit high transcriptional levels that remain stable in evolution. Highly-conserved regulatory elements active in most mammals also stabilise gene expression. Conversely, recently-evolved enhancers are typically weak, consistent with a largely neutral role in gene regulation, yet in large numbers can lead to gene expression divergence during evolution. These effects are consistently observed across the entire mammalian clade and robust to potential confounders, such as gene expression level and landscape complexity.

Overall, our results underscore how the evolutionary stability of gene expression is profoundly entwined with both the number and conservation of surrounding promoters and enhancers.

Expanded summary*:

Mammalian gene expression is controlled by collections of promoter and enhancer regions[1-3]. Numerous studies have documented the rapid evolution of mammalian regulatory elements, especially enhancers[4-8], and yet gene expression patterns are often highly stable between species[9-12]. How stable gene expression is maintained by rapidly evolving collections of enhancers and promoters is a fundamental question in evolutionary genetics.

To date, comparative approaches to gene regulation have largely focused on lineage-specific regulatory innovations, successfully identifying candidate regions driving lineage-specific gene expression patterns[4, 8, 13, 14] (reviewed in [15]). Our approach systematically extends these analyses to understand how evolutionarily plastic or stable regulatory elements influence gene expression in a representative tissue across the mammalian phylogeny. Evolutionarily conserved regulatory elements are thought to play a predominant role in gene regulation[16, 17]; the functional impact and broader role of regulatory elements with low signals of conservation has been the subject of speculation[18].

Here, we rigorously tested the contributions of both landscape complexity (number of regulatory elements) and conservation of regulatory activity on gene expression evolution, using an integrated dataset of promoter and enhancer histone marks and gene expression output from the same liver samples. Our methodology captures regulatory activities that range from essential to dispensable, and from highly-conserved across mammals to present in only one species (recently-evolved). This analysis revealed that gene expression levels and stability are reflected by the complexity of their regulatory landscape, both within a single species and across mammals. Additionally, regulatory activities highly-conserved in placental mammals exert a powerful stabilizing effect, associating with gene expression levels that are simultaneously high and evolutionarily stable. These discoveries extend previous reports connecting evolutionary constraint on promoter and enhancer activities with conserved expression outputs[4], and are consistent with the proposed functional relevance of evolutionarily constrained regulatory elements[17].

In contrast, recently-evolved enhancers contribute weakly to gene expression and transcriptional stability, consistent with a model whereby a sizable fraction of new-born enhancer elements have a neutral role on gene expression evolution[18]. However, we further have identified a set of genes that recurrently accumulate lineage-specific enhancers across species and that display increased expression divergence, indicating that genes with flexible expression levels better tolerate regulatory plasticity.

Our results underscore how the evolutionary stability of gene expression is profoundly entwined with both the number and conservation of surrounding promoters and enhancers. This effect is clear throughout our data, whether considering a full-scale,

reference-free map of all mammalian regulatory complexity, or investigating subsets of extremely conserved or divergent regulatory elements.

References:

1. Hay, D. et al. Genetic dissection of the alpha-globin super-enhancer in vivo. *Nat Genet* **48**, 895-903 (2016).
2. Shin, H.Y. et al. Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat Genet* **48**, 904-11 (2016).
3. Moorthy, S.D. et al. Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Res* **27**, 246-258 (2017).
4. Cotney, J. et al. The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* **154**, 185-96 (2013).
5. Xiao, S. et al. Comparative epigenomic annotation of regulatory DNA. *Cell* **149**, 1381-92 (2012).
6. Vierstra, J. et al. Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **346**, 1007-12 (2014).
7. Villar, D. et al. Enhancer evolution across 20 mammalian species. *Cell* **160**, 554-66 (2015).
8. Reilly, S.K. et al. Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155-9 (2015).
9. Brawand, D. et al. The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343-8 (2011).
10. Chan, E.T. et al. Conservation of core gene expression in vertebrate tissues. *J Biol* **8**, 33 (2009).
11. Merkin, J., Russell, C., Chen, P. & Burge, C.B. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338**, 1593-9 (2012).
12. Necsulea, A. & Kaessmann, H. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat Rev Genet* **15**, 734-48 (2014).
13. Pai, A.A., Bell, J.T., Marioni, J.C., Pritchard, J.K. & Gilad, Y. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet* **7**, e1001316 (2011).
14. Prescott, S.L. et al. Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68-83 (2015).
15. Reilly, S.K. & Noonan, J.P. Evolution of Gene Regulation in Humans. *Annu Rev Genomics Hum Genet* **17**, 45-67 (2016).
16. Panne, D. The enhanceosome. *Curr Opin Struct Biol* **18**, 236-42 (2008).
17. Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476-82 (2011).
18. Cooper, G.M. & Brown, C.D. Qualifying the relationship between sequence conservation and molecular function. *Genome Res* **18**, 201-5 (2008).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-270

Genome-wide patterns of regulatory divergence in adaptive pathways

Devjanee Swain-Lenz [1,*], Lisa Pfefferle [1], Gregory Wray [1]

[1]Biology, Duke University, Durham, United States

Abstract: A challenge in molecular evolution is to understand the genetic divergence underlying phenotypic divergence between species. Regulatory elements are thought to be the major source of phenotypic divergence, but we still lack a general understanding of how sequence divergence correlates to expression divergence. Primate adipocytes, cells that store lipids, are an exceptional model to study the relationship between genetic- and expression-divergence. To support the energy needs of our large brains, humans have adapted major changes in lipid metabolism compared to chimpanzees. Abundant evidence exists that human lipid metabolism is a positively selected pathway in response to brain size. For instance, as genes involved with lipid metabolism are upregulated in humans compared to chimpanzees. However, we do not know the causal mutations between humans and chimpanzees that mediate differences in lipid metabolism pathways. Using a high-throughput assay to identify regulatory elements, I will measure correlation of regulatory elements to gene expression to understand if species-specific elements are more likely to be associated with differential gene expression than elements shared between species. Additionally, I will analyze if species-specific elements are more likely to be under selection than elements shared between humans and chimpanzees and determine if transcription factor binding motifs predict species-specific regulatory elements. The results of this study will shed light on the relationship between chromatin structure, genetic sequence divergence and differential gene expression between species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-273

The driving force underlying the evolution of translational initiation context

Ke Li [1,*], Wenfeng Qian [1]

[1]Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Abstract: DNA sequences around the start codon AUG, also known as translational initiation contexts, play an important role in regulating protein level. However, the evolutionary pattern and driving force of these sequences remain unclear. To systematically examine the impact of translational initiation context, we performed high-throughput proteomics and ribosome profiling experiments in the budding yeast *Saccharomyces cerevisiae*, and calculated the similarity of the translational initiation context of each gene to the consensus sequence in the genome (initiation score hereafter). We found that the initiation score was strongly correlated with mRNA level rather than translational efficiency. Importantly, a similar pattern was also observed in mouse, fly, and *Arabidopsis*. We further generated 9300 variants of *GFP* with different translational initiation contexts, and found that translational initiation context regulated the mRNA level of *GFP*. Importantly, mutations that result inupstream AUGs (uAUG) usually reduced mRNA levels, and this impact was largely removed when the nonsense mediated decay pathway was shut down. Our study illustrates the regulatory role of translational initiation context on the mRNA level, demonstrated the coupling of translation and transcription, and shed the light on ultimately understanding the function and evolution of each nucleotide in the translational context.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-271

Coordinated stochastic expression variations of the GAL gene cluster in yeast

Haiqing Xu [1,*], Jianzhi Zhang [1]

[1]Ecology and Evolutionary Biology, University of Michigan, ANN ARBOR, United States

Abstract: There is mounting evidence that functionally related genes tend to be chromosomally clustered in eukaryotic genomes even after the exclusion of genes formed by tandem duplication, but little is known about the benefit of such clustering. We propose that, because neighboring genes are likely to be controlled by the same chromatin domain, the stochastic expression variations of neighboring genes among isogenic cells tend to be coordinated such that the expression ratio between them has smaller among-cell variations than that for unlinked genes. Consequently, gene clustering could be advantageous when the expression ratio of the clustered genes needs to be tightly regulated, for example, due to the accumulation of toxic compound when the expression ratio is misregulated. To test this hypothesis, we focus on the *GAL* gene cluster that includes *GAL1*, *GAL7*, and *GAL10*, which are chromosomally adjacent in the budding yeast *Saccharomyces cerevisiae*. The three *GAL* genes encode enzymes catalyzing consecutive reactions in galactose catabolism, with a cytotoxic intermediate metabolite. The yeast *GAL* cluster emerged through the relocations of originally unlinked genes in evolution. To quantify the potential benefit of the *GAL* gene clustering, we have used CRISPR/Cas9-based genome editing techniques to perturb the physical linkage between adjacent *GAL* genes followed by fitness essays of the mutant yeast. We are also measuring the among-cell fluctuation of the expression ratios between *GAL1* (or *GAL7*) and *GAL10* for linked and unlinked alleles. Data are being collected and analyzed at the moment, and we expect to report our findings at the SMBE meeting.

Expanded summary*: Position effect, which refers to the influence of the chromosomal location of a gene on its expression, is an essential aspect of gene expression regulation. Non-random genome configurations in eukaryotes offer profound materials for studying position effect. Among them, clustering of functionally related genes is one of the most intriguing patterns. It has been extensively documented that genes with related functions, such as genes encoding members of the same protein complexes or enzymes of the same metabolic pathways, tend to be linked chromosomally in eukaryotes, even after the exclusion of duplicated genes. Currently, two non-mutually exclusive hypotheses for explaining the benefits of gene clustering exist: genetic linkage and co-regulation. The genetic linkage hypothesis claims that clustering prevents co-adapted genes from separation by recombination. The co-regulation hypothesis asserts that divergently transcribed neighboring genes could be co-regulated by sharing a bidirectional promoter. Our hypothesis differs from the above latter hypothesis in that it involves the co-regulation of stochastic expressions of adjacent genes at the single-cell level rather than the co-regulation of mean expressions of a population of cells.

In our project, we study the effects of *GAL* genes clustering on co-expression noise in *S. cerevisiae*. First, we create a *cis*-deletion diploid mutant with only one intact copy of *GAL1* (or *GAL7*) and one intact copy of *GAL10* on the same chromosome, and a *trans*-deletion diploid mutant with only one intact copy of *GAL1* (or *GAL7*) and one intact copy of *GAL10* on different chromosomes. By competing between the *cis*- and *trans*-deletion mutants, we will determine whether the physical proximity of *GAL* genes enhances cell growth in galactose medium. Next, *GAL* genes will be fused with different florescence protein genes in a way that only one copy of *GAL1* (or *GAL7*) and one copy of *GAL10*, on the same or different chromosomes, will be tagged. Consequently, the ratios between single-cell gene expression levels of *GAL1* (or *GAL7*) and *GAL10* when they are physically linked or unlinked can be monitored by flow cytometry. Such data can provide mechanistic explanations for the potential fitness difference from the aforementioned competition experiment. Finally, we will disrupt the gene linkage by sequestering individual *GAL* genes with insulators, and anticipate that the differences in fitness and expression ratios of *GAL1*/*GAL10* (or *GAL1/GAL10*) between *cis*- and *trans*-constructs will diminish. Overall, our findings are expected to provide insights into the molecular basis of the potential benefit of gene clustering in eukaryotic genomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-292

Flip/flop mating-type switching in the methylotrophic yeast Ogataea polymorpha is regulated by an Efg1-Rme1-Ste12 pathway

Sara Hanson [1,*], Kevin Byrne [2], Kenneth Wolfe [2]

[1]Molecular Biology, Colorado College, Colorado Springs, United States, [2]School of Medicine and Medical Sciences, University College Dublin, Dublin, Ireland

Abstract: In *Ogataea (Hansenula) polymorpha* an environmental signal, nitrogen starvation, induces a reversible change in the physical structure of a chromosome. This process, mating-type switching, involves inverting a 19-kb DNA region to place either *MAT*a or *MAT*alpha genes under centromeric repression of transcription, depending on the orientation of the region. Here, we investigated the genetic pathway that controls switching. We first characterized the transcriptomes of haploid and diploid *O. polymorpha* by RNAseq in rich and nitrogen-deficient media, and found that there are no constitutively a-specific or alpha-specific genes other than the *MAT* genes themselves. We identified a switching defect in a related species (*O. parapolymorpha* strain DL-1) and mapped the defect by interspecies bulk segregant analysis to a frameshift in the transcription factor *EFG1*, which in *Candida albicans* regulates filamentous growth and white/opaque switching. By gene knockout, overexpression and ChIPseq experiments we established that *EFG1* regulates *RME1*, which in turn regulates *STE12*. All three genes are necessary for mating-type switching in response to nitrogen starvation, as well as for mating. Overexpression of *RME1* or *STE12* alone is sufficient to induce switching without a nitrogen signal. Our results show that the pathway controlling switching in *O. polymorpha* is substantially different from that in *S. cerevisiae*, which does not involve an environmental signal, and that it shares some components with mating-type switching in *Kluyveromyces lactis* and with white/opaque phenotypic switching in *C. albicans*. The downstream mechanism by which the chromosomal inversion occurs remains unidentified.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-296

Alu-derived alternative splicing events specific to Macaca lineages in CTSF gene

Sang-Je Park*, Young-Hyun Kim [1], Hyeon-Mu Cho [1], Jae-Won Huh [1], Sang-Rae Lee [1], Kyu-Tae Chang [1]

[1]National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology, Cheongju, Korea, Republic Of

Abstract: Cathepsin F, which is encoded by *CTSF*, is a cysteine proteinase ubiquitously expressed in several tissues. In a previous study, novel transcripts of the *CTSF* gene were identified in the crab-eating monkey deriving from the integration of an *Alu* element–*Alu*YRa1. The occurrence of *Alu*YRa1-derived alternative transcripts and the mechanism of exonization events in the *CTSF* gene of human, rhesus monkey, and crab-eating monkey were investigated using PCR and reverse transcription PCR on the genomic DNA and cDNA isolated from several tissues. Results demonstrated that *Alu*YRa1 was only integrated into the genome of *Macaca* species and this lineage-specific integration led to exonization events by producing a conserved 3' splice site. Six transcript variants (V1–V6) were generated by alternative splicing (AS) events, including intron retention and alternative 5' splice sites in the 5' and 3' flanking regions of *CTSF_Alu*YRa1. Among them, V3–V5 transcripts were ubiquitously expressed in all tissues of rhesus monkey and crab-eating monkey, whereas *Alu*YRa1-exonized V1 was dominantly expressed in the testis of the crab-eating monkey, and V2 was only expressed in the testis of the two monkeys. These five transcript variants also had different amino acid sequences in the C-terminal region of CTSF, as compared to reference sequences. Thus, species-specific *Alu*-derived exonization by lineage-specific integration of *Alu* elements and AS events seems to have played an important role during primate evolution by producing transcript variants and gene diversification.
Keywords: Primate, CTSF, Alu, Exonization, Alternative splicing

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-298

Bioinformatic Analysis of Expression and Regulation of Natural Antisense transcripts by Transposable elements in Human mRNA

Sugi Lee [1,*], Jaeeun Jung [1], Kunhyang Park [2], Dae-Soo Kim [3]

[1]University of Science and Technology, [2]Core Facility Management Research Center, [3]Rare Disease Research Center, KRIBB, Daejeon, Korea, Republic Of

Abstract: Recently, the importance of gene expression regulatory roles of natural antisense transcripts(NATs) is arising. It is well known that NATs are produced the opposite strand encoding a protein. Still, the precise transcriptional mechanisms initiated by NATs remain incompletely understood. And insertion of transposable element into the promoter region could affect the transcription of cellular genes. The main aim of this study was to find NATs in human genome that derived from the transposable elements. Using a bioinformatics approach, we searched the NATs using the human mRNA sequence from the UCSC Genome Browser. In this study, we established a set of very stringent criteria to identify precise NATs. We excluded unspliced mRNA sequence data to avoid possible contamination of genomic sequence. From our in silico analysis of human genome indicated that 1,079 NATs. There are 144 NATs identified to have been affected by transposable elements(TEs) during the cellular gene expression and 935 NATs without TEs. This result suggests that genes are regulated by NATs derived from the transposable elements in human genome. These findings may also provide understanding of the complex regulation mechanisms and dynamic evolutionary features during human evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-297

Identification and characterization of alternative splicing and polymorphic insertion in TSEN54 gene during primate evolution

Hyeon-Mu Cho*, Jae-Won Huh [1], Sang-Je Park [1], Young-Hyun Kim [1], Ja-Rang Lee [1], Se-Hee Choe [1], Sang-Rae Lee [1], Kyu-Tae Chang [1]

[1]National Primate Research Center, Korea Research Institute of Bioscience and Biotechnology, Cheongju, Korea, Republic Of

Abstract: *TSEN54* encodes a subunit of the tRNA splicing endonuclease complex, which catalyzes the identification and cleavage of introns from precursor tRNAs. Previously, we identified an *AluSx*-derived alternative transcript in *TSEN54* of cynomolgus monkey. Reverse transcription-polymerase chain reaction (RT-PCR) amplification and *TSEN54* sequence analysis of primate and human samples identified five novel alternative transcripts, including the *AluSx* exonized transcript. Additionally, we performed comparative expression analysis via RT-qPCR in various cynomolgus, rhesus monkey, and human tissues. RT-qPCR amplification revealed differential expression patterns. Furthermore, genomic PCR amplification and sequencing of primate and human DNA samples revealed that *AluSx* elements were integrated in human and all of the primate samples tested. Intriguingly, in langur genomic DNA, an additional *AluY* element was inserted into *AluSx* of intron eight of *TSEN54*. The new *AluY* element showed polymorphic insertion. Using standardized nomenclature for Alu repeats, the polymorphic *AluY* of the langur *TSEN54* was designated as being of the *AluYl17* subfamily. Our results suggest that integration of the *AluSx* element in *TSEN54* contributed to diversity in transcripts and induced lineage- or species-specific evolutionary events such as alternative splicing and polymorphic insertion during primate evolution.
Key words: *TESN54* gene; Alternative splicing; *Alu* element; Polymorphic insertion; Primate evolution

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-289

Comparison of X chromosome inactivation profiles across mammals suggests a role for gene function in determining inactivation status

Andrea J. Slavney*, Andrew G. Clark [1]

[1]Molecular Biology and Genetics, Cornell University, Ithaca, United States

Abstract: In mammals, X chromosome inactivation (XCI) facilitates dosage compensation between XX females and XY males. Some genes exhibit partial expression from the inactive X, and are termed XCI escapers. Cross-species comparisons of XCI profiles – i.e. which genes escape XCI or are completely inactivated – can help us understand whether XCI escape is primarily driven by inefficient XCI, or is positively favored for some biological function. To this end, we compared biological functions and evolutionary history of X genes in five species with known XCI profiles: human, mouse, opossum, horse, and dog. Where possible, we also examine XCI profiles derived from single-cell transcriptomics.

First, we compared XCI profiles from each species, and found limited conservation of XCI status across orthologs. Despite this, when we compared gene ontology annotations for XCI escapers and X-inactivated genes in each species, we observed that several biological processes were shared across species within each XCI category. Next, we compared signatures of selection between XCI escapers and X-inactivated genes within each species. In agreement with previous studies, we observed that human XCI escapers show a greater degree of conservation than X-inactivated genes. While this pattern generally held in other species, the magnitude of the difference between the XCI categories varied. These results suggest that despite XCI profiles being lineage-specific, they are shaped by similar selection pressures.

Expanded summary*: X chromosome inactivation (XCI) escape introduces gene expression variation between females and males, and variable XCI escape inflates variation among females. These differences may have important functional consequences. For instance, it has been suggested that some human and mouse XCI escapers perform female-specific functions, including some that are disrupted in X aneuploidy syndromes. However, it remains unclear whether XCI escape is generally a "bug" in the XCI system – that is, merely a consequence of inefficient XCI – or a functional "feature" necessary for normal biological function.

While divergence from the Y chromosome appears to drive evolution of XCI at the gene level, it is an imperfect predictor of XCI status: some genes with functional Y homologs fail to escape XCI, and many genes lacking functional Y homologs do escape. Physical and evolutionary drivers of XCI escape may vary across genes, with some expressed biallelically for functional reasons, and others escaping by based solely on their physical position. Comparing XCI profiles across mammals can provide insight into the ancestral XCI status of specific genes, as well as the approximate timing of changes in XCI status. With this information, we can identify genes for which XCI escape is more driven by selection, and therefore more relevant to human diesease.

At present, extensive XCI profiles are only available for three species: human, mouse, and opossum. Surprisingly, these show that XCI statuses of orthologous X genes are poorly conserved. For example, of over 160 human XCI escapers with mouse orthologs, only 13 also escape in mouse. But because the murine X shows large-scale structural rearrangements not seen in other mammals, including unusually small psuedoautosomal regions, this is not entirely unexpected. With this in mind, we use existing XCI profiles from human, mouse, and opossum, and novel horse and dog XCI profiles, in a comparative study of XCI evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-290

Unraveling the regulatory networks of maize leaf and seed development

Jinn-Jy Lin [1 2 3,*], Chun-Ping Yu [3], Wen-Hsiung Li [1 3 4]

[1]Institute of Molecular and Cellular Biology, National Tsing Hua University, Hsiuchu, [2]Bioinformatics Program, Taiwan International Graduate Program, Institute of Information Science, Academia Sinica, [3]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan, [4]Department of Ecology and Evolution, University of Chicago, Chicago, United States

Abstract: Maize (*Zea mays*) is an important crop and a model C4 plant.. The abundant genomic and transcriptomic data in maize greatly facilitate large-scale studies, such as gene expression profiles in different types of cells or tissues. However, the regulatory genomics in the maize genome is still not well studied. Here regulatory genomics refers to the study of relationships between functional elements and their regulators in a genome. The aim of this study is to identify maize transcription factors (TFs), transcription coregulators (TCs), transcription factor binding sites (TFBSs) and TF-target gene relationships.
We designed bioinformatics methods to dissect each aspect of the regulatory genomics in maize leaf and seed development. We focused on the construction of regulatory modules consisting of TF and TC genes and their target genes. We first updated the annotation of TF and TC genes in maize and foxtail millet (*Setaria italica*). We then predicted the TFBSs in the maize genome and studied the TF-TFBS relationships in maize leaf and seed development. Moreover, we developed a method to predict the target genes of maize TFs and inferred the regulatory networks in maize leaf and seed development. Our study provides an efficient workflow from characterization of TFs and TCs to their target genes and then the inference of the TF-target gene relationships in maize leaf and seed development.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG2

Studying cis-regulatory evolution with quantitative models of sequence function

Saurabh Sinha [1,*]

[1]Computer Science, University of Illinois, Urbana-Champaign, United States

Abstract: Several studies have examined the evolution of gene regulation by characterizing genome-wide changes in transcription factor (TF) binding across species, in some cases also relating these changes to divergence or conservation of gene expression. Other studies have investigated changes in regulatory activity between orthologous enhancers. Typically, these studies also explore whether changes in regulatory function are reflected in sequence changes, and are able to attribute functional change at least in part to divergence in transcription factor motif presence. However, the relationship between sequence divergence and regulatory evolution remains poorly understood, especially at a quantitative level. This is not surprising, given the general inability to decipher the cis-regulatory logic of enhancers.

We have in the past developed thermodynamics-based models of the sequence-to-function relationship and used these to accurately predict readouts of enhancers from their sequence, and to also predict genome-wide TF-DNA binding profiles. We believe that such models can significantly advance our understanding of cis-regulatory evolution at the sequence level, from qualitative (e.g., "are motif changes seen?") to quantitative (e.g., "how well are functional changes explained by sequence divergence?") I will present results from our recent work in this direction. In one study, we quantitatively examine the extent to which sequence and accessibility changes, taken separately or in combination, can predict TF occupancy divergence. In a second study, we analyze single nucleotide polymorphisms located within a developmental enhancer whose function we are able to model accurately, for evidence of compensatory mutations and avoidance of large effect SNPs.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-295

Genotype x Environment Interactions and the Maintenance of Polygenic Sex Determination

Richard Meisel [1,*]

[1]University of Houston, Houston, United States

Abstract: Sex determination (SD) evolves fast, with master regulators at the top of SD pathways often differing between species and even polymorphic within populations. Opposing selection pressures in males and females (intersexual antagonism) on sex determining variants or linked alleles has been proposed as the only mechanism for the stable maintenance of multiple master regulators of SD in a species (polygenic SD). However, spatially and temporally heterogenous environments can produce variable selection pressures that maintain other forms of genetic variation, yet heterogeneous selection has received little attention as a way to maintain polygenic SD. To address this limitation, we studied the housefly, which segregates for multiple male and female determining variants at different frequencies across natural populations. We used a population genetic model to identify fitness paramaters that explain the stable maintenance of polygenic SD. Genotype fitness is negatively correlated between males and females in our models, suggesting that sexually antagonistic selection is involved in the maintenance of polygenic SD. We also quantified the expression of a female-determining isoform in the SD pathway, and we found that misexpression in males depends on the male determining genotype and rearing temperature. The male genotype most common in northern populations (with colder temperatures) expresses the female isoform more when reared at higher temperatures. Our results provide evidence that genotype-by-environment interactions can affect the SD pathway, producing heterogenous sex-specific selection pressures that maintain polygenic SD.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-286

Transcriptional changes due to haploid selection

Roy Francis [1,*], Ghazal Alavioon [1], Simone Immler [1]

[1]Ecology and Genetics, Uppsala University, Uppsala, Sweden

Abstract: An inescapable consequence of sex in eukaryotes is the evolution of a biphasic life cycle with alternating diploid and haploid phases. The occurrence of selection during both phases has far reaching consequences for fundamental evolutionary processes including the rate of adaptation, extent of inbreeding depression and load of deleterious mutations, as well as for applied research into assisted fertilization. It has been a long-standing dogma that, unlike in plants, selection at the haploid gametic level in animals is of no great importance. Our research investigates the relative importance of selection at the gametic level for subsequent generations using the zebrafish *Danio rerio* as a model system. We find clear evidence for a substantial impact of selection on haploid sperm on offspring fitness, which is reflected both in offspring fitness from early life stages into adulthood as well as the adult transcriptome and genome. Our results indicate that selection on sperm has a global effect on the transcriptome. Up to 746 and 1047 significantly differentially expressed genes were identified in the Brain and Testes respectively. Gene ontology analyses reveal that the differentially expressed genes are involved in fundamental processes such as metabolic processes, signalling pathways, protein synthesis, morphogenesis and development. KEGG pathway analyses showed several upregulated genes in the oxidative phosphorylation and ribosome synthesis pathways.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-285

On the Mechanistic Nature of Epistasis in a Canonical cis-Regulatory Element

Mato Lagator [1,*], Tiago Paixao [1], Nicholas H. Barton [1], Jonathan P. Bollback [2], Calin C. Guet [1]

[1]IST Austria, Klosterneuburg, Austria, [2]University of Liverpool, Liverpool, United Kingdom

Abstract: Understanding the relation between genotype and phenotype remains a major challenge. The difficulty of predicting individual mutation effects, and particularly the epistatic interactions between them, has prevented the development of a comprehensive theory that links genotypic changes to their phenotypic effects. First, we experimentally explore the nature of epistasis in a canonical *cis*-regulatory element consisting of overlapping RNA polymerase and repressor binding sites. Then, we show that a general thermodynamic framework for gene regulation, which is based on a biophysical understanding of protein-DNA binding, accurately predicts the sign of epistasis and its environment-dependence. We show that both sign and magnitude of individual mutation effects are sufficient to predict the sign of epistasis and its dependence on the environment. Thus the thermodynamic model offers the correct null prediction for epistasis between mutations within and between DNA-binding sites. Our results indicate that a predictive theory for the effects of *cis*-regulatory mutations is possible from first principles, as long as knowledge about the essential molecular mechanisms and the constraints these impose on a biological system are accounted for.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-284

The Evolutionary Dynamics of Vertebrate Replication Origins and their Impact on Genome Architecture

Florian Massip [1,*], Marc Laurent [2], Caroline Brossas [2], Marie-Noëlle Prioleau [2], Franck Picard [1], Laurent Duret [1]

[1]université Lyon1, Villeurbanne, [2]Université Paris Diderot, Paris, France

Abstract: In vertebrate genomes, DNA replication is initiated at precise genomic regions, called replication origins (Oris). Despite the critical role of the replication process in maintaining genome integrity, the genetic features that induce the activity of Oris are not well understood. To decipher sequence features that are essential for Oris' activity, we conduct the first comparative analysis of replication origins in mammals and birds.
We detect Oris in chicken, generating the first genome-wide map of Oris in a bird, and compare it to human and mouse maps.
We find that despite a three-fold difference in genomes' size, Oris' density – one Ori every 30 to 50kb – is similar in mammals and birds. This suggests that the replication of genomes imposes constraints on Oris' density. In these three genomes, Oris are strongly associated to transcription start sites (TSSs), CpG islands (CGIs) and G-quadruplex (G4s). This confirms the importance of these elements in Oris' function.
Given the importance of replication in ensuring the inheritance of genetic information, one could expect Oris to evolve under strong purifying selection. Although 15% (8% resp.) of human Oris are homologous to mouse Oris (to chicken Oris resp.), three times more than what is expected by chance, we demonstrate that the conservation of TSSs and CGIs fully explains Oris' conservation. In addition, we find that Oris present similar sequence conservation than their close neighborhood.
Overall, our results suggest that replication origins undergo a rapid turnover and induce only mild constraints on the evolution of vertebrate genomes.

Expanded summary*: DNA replication follows a spatio-temporal program that ensures the faithful replication of genomes at each cell cycle. In vertebrate genomes, DNA replication initiates at precise genomic regions, called replication origins (Oris). Despite the critical role of the replication process in maintaining genome integrity, the genetic features inducing Oris' activity are not well understood. We propose the first investigation of Oris' evolutionary constraints in order to decipher sequence features essential for Oris' activity.
We map replication origins in chicken, by purifying and sequencing short nascent strands (SNS), thus generating the first genome-wide map of Oris in a bird. To compare chicken and mammalian Oris, we reanalyze published SNS sequencing data from human and mouse genomes. To ensure that the sensitivity of Ori detection is similar in all species, we sub-sample SNS reads to obtain identical sequencing depth in each species, and use the same peak-calling methodology to map Oris.
We detect 84250, 55000 and 25900 Oris in human, mouse and chicken respectively. These numbers are proportional to the size of their genome (3.10, 2.72 and 1.05 Gb respectively). Thus, despite a three-fold difference in genome size, the density in Oris – one per 30 to 50kb – is similar in mammals and birds, which suggests that replication imposes constraints on Oris' density. In these three genomes, Oris are associated to the same genomic elements, namely transcription start sites (TSSs), CpG islands (CGIs) and G-quadruplex (G4). This confirms the importance of these elements in Oris' function.
Given the importance of replication in ensuring the inheritance of genetic information, one could expect these regions to evolve under strong purifying selection. Surprisingly, we find that the constraints affecting Oris at the sequence level are rather weak. Although 15% (resp. 8%) of human Oris are homologous to mouse (resp. to chicken) Oris, which is three times higher than what is expected by chance, we demonstrate that the conservation of the genomic elements to which Oris are associated (TSSs , CGIs) fully explains Oris' conservation. In addition, we find that Oris present a level of sequence conservation similar to their neighborhood. Finally, TSSs, CGIs and G4s associated to Oris are not more conserved when they overlap an Ori.
Overall, our results suggest that replication origins undergo a rapid turnover and induce only mild constraints on the evolution of vertebrate genomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG14

How much does RNA-level regulation constrain exonic sequence evolution?

Rosina Savisaar [1,*], Laurence D. Hurst [1]

[1]The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom

Abstract: Exon evolution is constrained by the need to maintain interactions with splice factors, as well as potentially other RNA-binding proteins (RBPs). Is this pressure a major driver of mammalian exon evolution? We analysed enrichment and constraint on exonic motifs putatively recognized by RBPs. A subset of the motifs is indeed selectively maintained in coding sequence, while others, notably those recognized by intronic or UTR binders, are actively avoided. The extent of motif enrichment predicts the evolutionary conservation of the motifs. As constraint is also seen in intronless genes, not all RBP-related selection in exons is splice-associated. Overall, we estimate in two independent nucleotide-controlled comparisons (mouse-rat and human-macaque) that the net effect of motif preservation is modest, with an ~2-3% reduction in the synonymous rate of evolution of the median gene. While this figure agrees with estimates for constraint imposed by exonic splice enhancers, it clashes strikingly with experimental splice reporter studies, which have found a large proportion (up to ~60%) of exonic mutations to disrupt splicing. Why do the two approaches come to such different conclusions? Among other factors, we suggest that the cell's ability to buffer changes in relative dosage of splice isoforms is a leading explanation.

Expanded summary*: In addition to specifying protein structure, coding exons have to maintain interactions with splice factors and other RNA-binding proteins (RBPs). That such overlapping signals exist within coding sequence (CDS) is well-established but the scale of these phenomena is unclear. Notably, are functional (i.e. selectively maintained) exonic splice regulatory elements rare, occurring mainly at a few well-defined positions in hard-to-splice exons? Or are our exons filled to the brim with various splice control signals, whose configuration needs to be preserved for correct splicing? Solving this problem will give a clearer picture of the forces driving exon evolution and will hint at the extent to which alternative splice patterns are functional. It will also help predict how frequently coding mutations might lead to potentially disease-causing splice alterations.

To tackle this issue, I have studied how *k*-mers putatively recognized by RBPs evolve in CDSs. Most strikingly, I have discovered that our CDSs are constrained not only to maintain necessary RBP interactions but also to avoid inappropriate ones: motifs associated to CDS-binding RBPs tend to be enriched and conserved over expected in CDS, whilst there seems to be active selection against gaining many of the motifs putatively recognized by intronic or UTR binders. These conclusions are also valid for intronless genes, implying that splice regulation is not the only source of RBP-associated constraint on exonic regions.

More quantitatively, in both human and mouse, it appears that the need to preserve RBP target motifs reduces by ~2-3% the synonymous rate of evolution of the median gene. Selectively maintained RBP target sites would therefore either be rare or, if frequent, then only weakly constrained. Although experimental studies have found a large proportion of exonic sites to disrupt splicing when mutated, the evolutionary analysis suggests that these effects are often not visible to selection.

Finally, how does the constraint distribute across elements? Is the conservation limited to a few sites, which are under strong negative selection, or is there also a large class of sites under weak negative selection? To find out, I am currently determining the distribution of fitness effects for incoming mutations at exonic splice enhancers and within other *k*-mers putatively recognized by RBPs. The results will take us closer to a detailed understanding of how non-coding function shapes exon evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-288

Social integration influences mitochondrial DNA regulation in rhesus macaques

Reena Debray [1,*], Jordan Kohn [2], Mark Wilson [2], Luis Barreiro [3], Noah Snyder-Mackler [1], Jenny Tung [1]

[1]Duke University, Durham, [2]Emory University, Atlanta, United States, [3]University of Montreal, Montreal, Canada

Abstract: In humans and other social mammals, social subordinacy and social isolation have been shown to predict shorter lifespans and higher rates of disease, and thus can have major consequences for individual fitness. These effects are thought to be driven in part by chronic social stress, but their molecular underpinnings are not well understood. Here, we tested the hypothesis that changes in the mitochondrial genome regulation could be involved. To do so, we investigated the effects of experimentally manipulated dominance rank (a source of chronic social stress) on mitochondrial DNA copy number (mtDNA copy number) and heteroplasmy in captive female rhesus macaques (*Macaca mulatta*; n=45). We quantified mtDNA copy number in five purified white blood cell populations using qPCR, and assessed heteroplasmy using RNA-seq. We found that dominance rank did not predict either mtDNA copy number or heteroplasmy rates. In contrast, macaques that groomed more had higher mtDNA copy number (ß=0.039, p=0.014), in B cells, monocytes, and cytotoxic T cells. However, grooming frequency was not a significant predictor of the number of heteroplasmic sites in a sample (ß=0.002, p=0.884). Together, these results suggest a previously unappreciated contribution of mtDNA copy number in immune cells to the relationship between social connectedness and individual fitness. These selective pressures on social behavior may have contributed to the evolution of social groups in humans and other social mammals.

Statement: I am currently in my junior year at Duke University. I am majoring in biology with a genomics concentration. This fall, I will be applying to Ph.D. programs in evolutionary biology and genomics. My interest in an academic career is motivated by my research at Duke, which has focused on animal behavior. However, I am interested in changing my research focus in graduate school. I am interested in studying how aspects of genome structure, such as mating system and ploidy, influence adaptation, diversification, and extinction rates. This is an important area in a time of rapid ecological change, and I hope to conduct research that can inform conservation efforts.
The SMBE 2017 annual meeting provides an ideal opportunity for me to learn about methods, current questions, and potential advisors in this field, before I apply to graduate programs in the fall. Many of the symposia align closely with my research interests, including Evolutionary Genomics of Structural Variation, Polyploidy and Hybridization, and Genomic Mechanisms of Speciation. Presenting a poster at SMBE 2017 would be a valuable experience for me. I have presented at poster sessions before, but I have never attended a meeting focused specifically on molecular biology and evolution. I found that talking with others about my poster is a great way to spark new ideas about next steps. SMBE 2017 would provide an opportunity for me to develop my plans for my senior thesis by talking with researchers in similar fields.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG12

Evolution of transcriptional and translational regulation during adaptation

Premal Shah*

Abstract: Studies of how organisms evolve in the lab have identified several key features of the evolutionary processes such as the dynamics of clonal interference and epistatic interactions between adaptive mutations. While the role of individual mutations on organismal fitness have been characterized, how these mutations influence faster growth remains a critical gap in our understanding of evolutionary dynamics. Here we provide a mechanistic understanding of phenotypic changes that occur at the transcriptional and translational level during adaptive evolution.
Using the long-term evolution experiment (LTEE) in *E. coli* as a model system, we have generated RNA-seq and ribosome profiling datasets in the two ancestral and 12 evolved lineages from 50,000 generation. Surprisingly, we find that transcriptional changes over 50,000 generations are only weakly correlated (R=0.08–0.4) with translational changes. Furthermore, most transcriptional changes tend to be buffered either partially or entirely at the translational level. Interestingly, we find that genes that are differentially regulated across all replicate lineages have lower translation efficiencies in the evolved lines relative to ancestral strains, while differentially regulated genes unique to each lineage have higher translation efficiencies in evolved lines. As a result, this approach provides a deeper understanding of the mechanistic basis of adaptive evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG1

Genetic models of human evolution

James Noonan*

Abstract: Uniquely human traits, such as the increased size and complexity of our brain, are encoded within the millions of genetic changes that distinguish us from other primates. Some of these changes, such as those located in Human Accelerated Regions (HARs), have been tentatively linked to human phenotypes. However, the precise role of these human-specific genomic changes in human traits remains elusive. This is because we lack two essential tools: the means to overcome the species barrier and employ the power of experimental genetics to study uniquely human genomic features in model organisms, and the ability to access and compare developmental processes in humans and other great apes. We will describe our recent work combining reverse genetic models of human-specific sequence changes with comparative analyses of primate development to identify the biological pathways, mechanisms, and cell types that were altered in our evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG9

The evolution of mammalian gene expression across multiple dimensions

Henrik Kaessmann [1], Margarida Cardoso Moreira[*]

[1]Center for Molecular Biology, Heidelberg University, Heidelberg, Germany

Abstract: A major goal in biology is to understand the molecular basis of phenotypic evolution, in particular that of humans and other mammals. Gene expression changes, due to regulatory mutations, are thought to underlie much of phenotypic innovation. Transcriptome studies from us and others based on RNA sequencing (RNA-seq) data for adult organs previously provided initial insights into mammalian gene expression evolution and its phenotypic implications. However, phenotypic evolution is thought to be largely founded on developmental regulatory changes, which determine species-specific tissue morphologies and thus lay the foundation for their typical physiological properties. We therefore added the developmental dimension to our endeavors by generating and analyzing RNA-seq data for more than 2,500 pre- and post-natal developmental samples (ranging from early organogenesis to adulthood) from various organs (cortex, cerebellum, heart, kidney, lover, testis/ovary, placenta/decidua) across representative mammals and an avian outgroup (human, macaque, mouse, rat, rabbit, opossum, red jungle fowl). Based on these data, we have analyzed the evolutionary dynamics of gene expression and phenotypic ramifications for various major aspects of the transcriptome in several projects that include protein-coding gene expression levels, alternative splicing, long noncoding RNAs, microRNAs and sex-biased expression. In a second main line of research, we moved to the next principal regulatory layer of protein-coding gene expression – translation – by producing extensive ribosome profiling and matched RNA-seq data for three major organs (cortex, liver, testis) from five mammals and a bird (human, macaque, mouse, opossum, platypus, red jungle fowl). Our analyses of these data revealed various intriguing facets of protein synthesis evolution, including patterns of compensatory change (relative to transcription), adaptive changes driven by positive selection, and regulatory mechanisms underlying evolutionary shifts in translation rates. Overall, our new projects provide fundamental novel insights into gene expression evolution and its contribution to the specific organ biology of different mammals, including that of our own species. I will present selected highlights of our recent endeavors.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG3

The effects of background selection in the human genome reveal extensive purifying selection on non-coding regions

David Murphy [1,*], Guy Sella [1]

[1]Biological Sciences, Columbia University, New York, United States

Abstract: Recent studies of the effects of linked selection have made significantly better quantitative predictions about genome-wide diversity levels in humans and other species (e.g., McVicker et al., 2009, Elyashiv et al., 2016). Here we show that a model of background selection can accurately predict diversity levels throughout the human genome, explaining more than 40% of the variance on the 1Mb scale. The model also predicts fine scale features of the data, such as the depth and width of the trough in mean diversity within 0.05cM of coding regions. But perhaps even more interesting is what our inferences tell us about purifying selection in the human genome. Notably, they suggest that more than 80% of deleterious mutations occur in non-coding regions and that the best predictor of these regions is phylogenetic conservation in primates. Our results further suggest that we are missing many recent selection targets that arose in humans since the common ancestor of apes and that current functional genomic annotations contribute negligibly to predicting these regions. We estimate that ~46-92% of mutations in selected regions are deleterious, where the range reflects uncertainty about the fraction of selection targets missing from our model. Finally, our results suggest that background selection is the dominant mode of linked selection in humans and that it substantially affects diversity levels in most of the genome.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-258

The Evolution of Phenotypically Invariant Genetic Networks

Joshua Schiffman [1,*], Peter Ralph [2 3]

[1]Molecular and Computational Biology, University of Southern California, Los Angeles, [2]Institute of Ecology and Evolution, [3]Department of Mathematics, University of Oregon, Eugene, United States

Abstract: I will outline an analytical theory to study the evolution of biological systems such as gene regulatory networks, borrowing insight and tools from control engineering, systems identification, and dynamical systems theory. I will describe a null model of regulatory network evolution by analytically describing the set of all linear gene networks (of any size) that produce identical phenotypes -- and the evolutionary paths connecting them. In the idealized case of a perfectly adapted population, constant selection, and a static environment, we observe neutral evolution as a random walk over the phenotypically-invariant network-space. Under neutral conditions, this model can provide descriptions of expected network size and connectivity under mutation-selection equilibrium, estimate the rate of regulatory rewiring, and the rates at which Dobzhansky-Muller incompatibilities arise in reproductively isolated populations. This analysis provides insight into the mechanisms and parameters important for understanding developmental systems drift, network rewiring, evolvability, epistasis, and speciation, as well as the tenuous connection between network architecture and function.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG6

Detecting stabilizing selection on gene expression level from human genetic and expression variation data

Ziyue Gao [1 2,*], Emily Glassberg [3], Jonathan Pritchard [1 2]

[1]Department of Genetics, Stanford University, [2]Howard Hughes Medical Institute, [3]Department of Biology, Stanford University, Stanford, United States

Abstract: Comparison of gene expression levels across species suggests that gene expression is constrained over evolutionary timescales, implying purifying selection on mutations that modify expression levels. Yet, recent transcriptomic studies have uncovered large amounts of heritable expression variation and standing regulatory polymorphism in human populations. Here we explore the relative importance of mutation, drift, and selection in shaping the human expression variation. In particular, we aim to characterize the prevalence and strength of selection on expression levels. We leverage human diversity data from large-scale transcriptome-sequencing projects such as DGN and GTEx to perform analyses of expression quantitative trait loci (eQTL) and allele-specific expression (ASE). The joint distribution of allele frequency and effect size of identified eQTLs hints at weak selection against expression-altering variants and varying strengths of constraint across groups of genes. However, eQTL-based analyses are limited by the low power of detecting rare or small-effect variation and inaccurate estimation of effect size due to linkage disequilibrium between multiple regulatory variants. In contrast, ASE assays minimize the influences of environmental and/or technical noise and enable us to evaluate the impacts of rare cis-variants. By combining the GTEx RNA-seq and whole-genome sequencing data, we detect a robust genome-wide trend that genes with more upstream heterozygous sites tend to show stronger ASE bias; importantly, on average, a rare variant contributes significantly more than a common variant to this bias. Finally, we observe intriguing heterogeneity across gene groups and tissue types. Together, our analyses of human genetic and expression variation patterns provide clear evidence for prevalent and heterogeneous stabilizing selection on gene expression levels, even over short timescales.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-291

Evolution of the PIWI/Argonaute gene family in vertebrates

Amanda Black [1,*], Michael Vandewege [2], Federico Hoffmann [1]

[1]Biochemistry and Molecular Biology, Mississippi State University, Mississippi State, [2]Biological Sciences, Texas Tech University, Lubbock, United States

Abstract: The genes in the Argonaute superfamily encode for a structurally conserved set of proteins that interact with small non-coding RNAs and play key roles in multiple RNA interference (RNAi) pathways. The PIWI/Argonaute proteins bind to smalls RNA to form RNA-induced silencing complexes (RISC), which identify targets through sequence complementarity and silence them via methylation or degradation, and because of this, the proteins play key roles in transcription and post-transcriptional regulation of gene expression. From a sequence standpoint, these genes are classified into two groups: the Argonaute proteins, which interact with small interfering RNAs (siRNAs) and microRNAs (miRNAs), and the PIWI (P element-induced wimpy testis) proteins, which associate with PIWI-interacting RNAs (piRNAs). There are eight members of this family in most mammals, four in each group (AGO1-4 and PIWIL1-4), and although the vertebrate Argonaute and PIWI proteins form clear monophyletic groups, relationships within these two subgroups are not clear nor is their evolutionary history or phyletic distribution. In the current study, we integrate phylogenetic and synteny analyses to reconstruct the duplicative history of the PIWI/Argonaute gene family of vertebrates. Our results indicate that all the vertebrate Argonautes and PIWI emerged prior to the split between cartilaginous and bony fish and identify key differences in this regard between jawless and jawed vertebrates. Although only found in mammals, our phylogenies indicate that PIWIL3 is a relatively old paralog that has been lost multiple times. We discuss the potential implications of variation in gene content and expression patterns.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

POA-300

Genome-wide enhancer - target gene regulatory maps in two vertebrate genomes

Yves Clement [1,*], Hugues Roest Crollius [1]

[1]Institut de Biologie de l'ENS, Paris, France

Abstract: Enhancers are regulatory regions which drive the fine-scale regulation of gene expression among tissues and life stages. Because they can regulate genes at a long distance (>100 kbps), identifying their target genes is challenging.
We previously developed a method to identify evolutionary conserved target genes of enhancers using comparative genomics in vertebrates. Here, we applied this method to human and zebrafish in two separate analyses and found ~50,000 enhancers targeting ~17,000 genes in zebrafish and ~1,300,000 enhancers targeting ~18,000 genes in human. Among these, the core set of human-zebrafish conserved target genes outlines genes regulated by enhancers dating back to the ancestral vertebrate, and which are associated with development and brain functions. Moreover, we find that enhancer – target gene interaction distances scale remarkably well with genome size, suggesting the absence of physical constraints associated with physical distances.
We studied enhancer evolution following gene duplication and found that conservation of enhancers between duplicated genes decreases gradually with time, while the combined number of enhancers increases with time. Remarkably, both patterns are consistent with the evolution of gene expression profiles. Finally, both conservation of ancestral enhancers and acquisition of new enhancers are biased towards one gene in a duplicated pair, showing that genes are not equal within duplicated pairs.
Together, these results provide key insights into the mechanisms and evolution of gene regulation in vertebrates.

Expanded summary*: Long range regulatory regions (or enhancers) are regions that fine-tune gene expression during time and in tissues and cell types. While identifying these enhancers along genomes is possible through a variety of methods (e.g. sequence evolution, epigenetics marks), finding which genes they regulate is more challenging. Most methods either over-simplify enhancer - target gene relationships, leading to false predictions, or rely on costly experimental procedures, thus only applicable in a handful of organisms. We previously developed a method to 1) identify putative enhancers in genomes by looking for conserved regions in multiple alignments and 2) identify their target genes by looking for conserved enhancer - target gene associations in multiple species, which we measure by computing association scores. In this approach, the rationale is that the enhancer - target gene link, if functional, will be conserved by natural selection.
We applied this method to human and zebrafish in two separate analyses. Zebrafish is an ideal species as it is a model organism for vertebrate development, during which many enhancers are known to act in a time and tissue specific way. We were able to identify ~50,000 enhancers targeting ~17,000 genes in zebrafish and ~1,300,000 enhancers targeting ~18,000 genes in human. In these putative enhancers, coverage by functional features like histone marks (especially marks associated with development in zebrafish and with embryonic stem cells in human) increases with association scores, indicating our method is able to predict functional enhancer - target gene interactions. We are further validating our predictions with CRISPR-Cas9 experiments to show that our method correctly predicts functional interactions.
We then sought to find core set of genes with conserved regulation in vertebrates. We thus compared enhancer target gene interactions between human and zebrafish and found a set of conserved interactions involving ~1,500 enhancers and ~450 genes in both species, with genes expressed predominantly during development and in brain. This highlights functions for conserved gene regulation in vertebrates. We also found that enhancer - target gene interaction distances scales remarkably well with genome size, suggesting the absence of physical constraints associated with physical distances.
Finally, these predictions allow us to study the evolution of enhancers following the duplication of their target genes. We found that, while the conservation of regulatory regions between gene duplicates decreases gradually and rapidly over time, their combined number of enhancers increases with time. Showing acquisition of new enhancers after duplication. Interestingly, both patterns of evolution are consistent with patterns of gene expression evolution. We also found that gene duplicates are not equal within pairs: after duplication, both conservation of ancestral enhancers and acquisition of new enhancers is biased towards one gene.
Globally, we provide genome-wide enhancer and target gene predictions in two vertebrate genomes. We used these predictions to gain key insights on mechanisms of regulation in these genomes, and on the evolution of gene regulation in vertebrates.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolution of gene regulation

OT-EOG5

Cnidarian microRNAs shed new light on the evolution of post-transcriptional regulation of gene expression

Maayan Agron [1], Arie Fridrich [1], Vengamanaidu Modepalli [1], Reuven Aharoni [1], Yehu Moran [1,*]

[1]Ecology, Evolution and Behavior, The Hebrew University of Jerusalem, Jerusalem, Israel

Abstract: Over the past decade small RNAs such as microRNAs (miRNAs) have been shown to play pivotal roles as post-transcriptional regulators of gene expression in various physiological processes in plants and animals. miRNAs can be found in a wide range of animals, yet their functions were studied almost exclusively in members of the Bilateria such as arthropods, nematodes and vertebrates. Thus, studying their function in non-bilaterian phyla as Cnidaria (sea anemones, corals, hydras and jellyfish) is crucial for understanding the evolution of miRNAs in animals and can provide important insights into their roles in the last common ancestor of Cnidaria and Bilateria that lived more than 600 million years ago. Our results indicate that miRNAs in Cnidaria frequently have a nearly perfect match to their messenger RNA targets, resulting in target cleavage (slicing). This mode of action is common to plant miRNAs, but very rare in Bilateria. Further, we show that like plant miRNAs, most cnidarian miRNAs carry a 2-O'-methyl group that stabilizes them during full complementarity annealing to their target. These findings together with our discovery of cnidarian homologs of HYL1, a protein involved in miRNA biogenesis in plants, raises the intriguing possibility that the miRNA pathway existed in the common ancestor of plants and animals and puts in question the commonly suggested scenario of an independent emergence of the miRNA pathway in animals and plants.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

OTH-EG3

The earliest maize from Tehuacan Mexico: paleogenomic evidence of gradual domestication in a 5,300 year old crop.

Miguel Andrés Vallebueno Estrada*, Isaac Rodriguez-Arevalo [1], Sara Garcia-Morales [1], Javier Martinez Gonzalez [2], Angel Garcia Cook [3], Jean-Philippe Vielle-Calzada [1], Rafael Montiel [1]

[1]UGA LANGEBIO, CINVESTAV, Irapuato, [2]Salvamento Arqueologico, [3]Estudios Arqueologicos, INAH, Mexico DF, Mexico

Abstract: A long history of archaeobotanical records indicates that the Tehuacan Valley in Mexico was an important center of early Mesoamerican agriculture, providing evidence of the complexity of maize domestication over the last 5,000 years. Owing to expeditions conducted by MacNeish and his team in the early 1960s, we conducted a new exploration of several rockshelters located in different regions of the Tehuacan valley, uncovering more than 100 nonmanipulated maize specimens dating from 5,300 to 1,000 years before present (BP), and allowing genetic comparisons that incorporate both the temporal and geographical scale to the perspective of ancient maize evolution. Our initial studies show that the earliest maize from San Marcos was a partial domesticate diverging from the landraces and containing ancestral allelic variants that are absent from extant maize populations across the genome, particularly at several loci important for domestication. The genomic comparison of three temporally convergent 5300-5000 BP samples indicated that they were unusually homozygous and genetically similar, suggesting the earliest maize from San Marcos was already inbred. The *de novo* assembly of their genome revealed unforeseen structural re-arrangements as well as insertion-deletion and gene variants that are likely to be absent from extant populations. We hypothesize that this structural variation could be related to environmental adaptations or to the size and structure of ancient maize populations. Our studies open new perspectives for discovering past genetic variability that could be useful for present or future agricultural conditions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

POB-343

Ancient domestication of a viral envelope gene in the mosquito lineage

Katie Kistler*, Lisa Kursel, Frances Welch, Adriana Ludwig, Harmit Malik

Abstract: Host-pathogen interactions drive evolution, selecting for modifications to existing genetic elements that confer a competitive advantage. Occasionally, hosts selectively retain integrated pathogen genes, co-opting them for novel host functions. In the context of an evolutionary struggle for dominance between host and pathogen, an especially striking example of this retention is host domestication of viral envelope (*env*) genes, the very genes that enable a virus to enter the host cell. The *syncytin* genes of mammals represent an example of such host domestication, in which distinct *env* genes have been independently adopted by different mammalian lineages for placental function. *Env* domestications have also been observed in invertebrates, including the *Iris* gene in *Drosophila* species; however, their function is unknown. Here, we report the identification of a domesticated viral *env* gene in the Anophelinae and Culicinae mosquito lineages. We estimate that this *Iris-like* gene was acquired at least 100 million years ago, making it one of the most ancient *env* domestications known to date. Its highly conserved open reading frame and high expression in the adult mosquito indicate that *Iris-like* may be functional and under purifying selection. *Iris-like* is predicted to retain the viral characteristic of a polyprotein that is cleaved into two peptides. Conservation of this cleavage points toward potential functions in membrane fusogenicity, receptor recognition or immunosuppression. Additionally, we have detected domesticated *env*s in other insects such as moths, ants and aphids, suggesting that such domestication events may be common in arthropods, and a potential source of gene novelty.

Expanded summary*: Mosquitoes are the primary transmission vectors for the pathogens that cause malaria, dengue fever, West Nile fever and many other diseases. These debilitating, and sometimes lethal, viruses and parasites are transmitted to hundreds of millions of humans each year. During my two-year tenure as a technician in the laboratory of Leslie Vosshall at Rockefeller University, we sought to characterize the sensory pathways of the mosquito *Aedes aegypti* that enable its specialized behaviors and contribute to its potency as a vector for human diseases. To achieve this, we established CRISPR-Cas9 in this species and coupled genetic knock-outs with quantitative behavioral assays to determine how specific sensory receptors enable *Ae. aegypti* to accurately detect and locate human hosts and appropriate egg-laying locations.

As a first-year graduate student in the laboratory of Harmit Malik at the University of Washington, I am excited to try to understand mosquito genetics and the host-parasite interaction from a different angle. Evolutionary analysis can be a powerful method to identify and interrogate important genes and biological functions. This potentially includes crucial genetic elements that have contributed to the mosquito's specialized ability to harbor many human pathogens at little or no cost to the mosquito and to integrate sensory cues to seek out human hosts.

The laboratory of Harmit Malik specializes in gaining new insights into host-parasite biology through evolutionary approaches. Host-pathogen interactions drive evolution, selecting for modifications to existing genetic elements that confer a competitive advantage. Occasionally, hosts selectively retain integrated pathogen genes, co-opting them for novel host functions. By poaching genes from pathogens, hosts are not just engaging in an arms race, but are actually gaining an advantage from an otherwise antagonistic interaction.

In the context of an evolutionary struggle for dominance between host and pathogen, an especially striking example is host domestication of viral envelope (*env*) genes, the very genes that enable a virus to enter the host cell. The *syncytin* genes of mammals represent an example of such host domestication, in which distinct *env* genes have been independently adopted by different mammalian lineages. In this way, mammals have usurped the fusogenic and immunosuppressive protein functions of viral genes for placental formation and function.

Although rarer, *env* domestications have also been observed in invertebrates, including the *Iris* gene in *Drosophila* species; however, their function is unknown. Here, we report the identification of a domesticated viral *env* gene in the Anophelinae and Culicinae mosquito lineages. We estimate that this *Iris-like* gene was acquired at least 100 million years ago, making it one of the most ancient *env* domestications known to date. Its highly conserved open reading frame and high expression in the adult mosquito indicate that *Iris-like* may be functional and under purifying selection to preserve a crucial function. *Iris-like* is predicted to retain the viral

characteristic of a polyprotein that is cleaved into two peptides. Conservation of this cleavage site points toward potential functions in membrane fusogenicity, receptor recognition or immunosuppression. Additionally, we have detected domesticated *env*s in other insects such as moths, ants and aphids, suggesting that such domestication events may be common in arthropods, and a potential source of gene novelty.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

OTH-EG5

Adaptation to nine thousand years of diet in Asia

Srilakshmi Raj [1,*], Matthieu Foll [2], Allison Pei [3], Laurent Excoffier [4], Dorian Fuller [5], Toomas Kivisild [6], Andrew Clark [1]

[1]Molecular Biology and Genetics, Cornell University, Ithaca, NY, United States, [2]International Agency for Research on Cancer, Lyon, France, [3]Cornell University, Ithaca, NY, United States, [4]Population Genetics, Institute of Ecology and Evolution, Bern, Switzerland, [5]Institute of Archaeology, University College London, London, [6]Biological Anthropology, University of Cambridge, Cambridge, United Kingdom

Abstract: Agriculture and domestication created dietary changes that influenced human genomic variation. Shifts creating the strongest impacts have been explored extensively. The process of agriculture and domestication, however, was complex and spread over several thousands of years across different global origins. We hypothesize that the diet shifts introduced by plant and animal domestication must have rendered a greater impact on the human genome than previously explored, and these impacts varied among human populations. We combined detailed archaeological evidence of domestication with modern-day dietary information to reconstruct high resolution models of diet across Asian populations for the last 9000 years, incorporating mode of subsistence, food preparation techniques, time since domestication of the crop, micro- and macro-nutrient composition, and percent of diet. We used three Bayesian methods, BayeScan, BayeScEnv, and Baypass, to identify correlations between common genetic variants in 29 modern-day Asian populations and 9 dietary variables. We found an enrichment for genetic pathways associated with salivary gland morphology, insulin secretion, taste perception, olfaction and kidney development in the top 1% of gene regions correlated with our dietary variables. Many of these genes and gene families have not been previously reported to be under selection, especially due to diet.

We present a case for archaeobotanical evidence as a powerful tool for understanding how historical human niche construction influenced modern human genetic variation. As our knowledge of the timing and spread of agricultural domestication increases, we can use similar techniques to more accurately measure the impact of subtle environmental changes on the human genome.

Expanded summary*: One of the most dramatic animal domestication events occurred some 8000 years ago when humans domesticated themselves to a diet of domesticated plants and animals. The major changes in the genome resulting from the shift from a hunter-gatherer lifestyle to an agrarian one suggest that 1) dramatic and sharp lifestyle changes resulted in equally dramatic changes in the genome, and 2) many of these changes resulted specifically from sharp dietary changes. Milk-drinking cultures have a markedly higher frequency of *LCT* polymorphisms compared with non-milk-drinking cultures, and *ADH1B* polymorphism frequencies correlate with time of rice domestication [1,2]. Studies of ancient human DNA from 8000 years ago also suggest that European genomes may have undergone selection for loci involved in decreased height, in vitamin D levels, in fatty acid metabolism, in pigmentation and in hair thickness over this time period [3].

These and other studies on the relationship between domestication, diet and human evolution have used what is known about current dietary habits and lifestyles of populations, and focused on dramatic changes that influenced the genome in selective sweeps. In contrast to these studies, I treated diet as a dynamic variable that changed in composition and through time, using high-resolution archaeological data to construct the model. I worked with a nutritionist to identify the optimal way to combine the dietary variables in terms of

glycemic index and glycemic load. I also quantified diet through its macronutrient and micronutrient composition, to understand how dietary changes influenced nutrition. I also used three different, but complementary, statistical techniques to measure the correlation between these subtle environmental changes and modern genetic variation.

Construction of the dietary model and application to genetic data was exhaustive and required a truly interdisciplinary collaboration and approach. This work showed correlations between long-term dietary habits and gene ontologies associated with salivary gland development, insulin signaling, kidney development, and sensory perception. Many of these associations were found in pathways that have never before been suggested to evolve due to domestication. This work also demonstrates that archaeological data can be successfully combined with nutritional and genomic data to better understand how humans evolved in a changing environment.

1. Tishkoff, S. A. *et al.* Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* 39, 31–40 (2007).
2. Peng, Y. *et al.* The ADH1B Arg47His polymorphism in east Asian populations and expansion of rice domestication in history. *BMC Evol. Biol.* 10, 15 (2010).
3. Mathieson, I. *et al.* Eight thousand years of natural selection in Europe. *bioRxiv* 016477 (2015). doi:10.1101/016477

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

POB-347

Demographic modelling of rabbit genomes fails to recover historically known population history

Evan Irving-Pease [1,*], Laurent Frantz [1], Greger Larson [1]

[1]The Palaeogenomics and Bio-Archaeology Research Network, Research Laboratory for Archaeology and History of Art, University of Oxford, Oxford, United Kingdom

Abstract: Rabbits were domesticated during the historical era from an extant wild population, indigenous to the Iberian Peninsula and the south of France. According to historical records, rabbits are presumed to have been domesticated by monks ~ 600AD and to have undergone a severe domestication bottleneck. Rabbits are therefore an ideal model to assess the reliability of population genetics techniques to model the process of domestication. The demographic history of wild rabbits is also well characterised, since they were deliberately dispersed outside of their natural range, and their expanded populations crashed by at least 90% in the early 1950s due to the myxomatosis epizootic. These population expansions and crashes in the domestic and wild populations should be visible within the genomes of modern individuals.

Here we used models based upon the joint site frequency spectrum to test the ability of these methods to infer the reported population histories. Our results firstly demonstrate that a simple isolation-with-migration (IM) model can achieve a good fit, despite the omission of a myxomatosis bottleneck. Secondly, the coalescence time between modern wild and domestic rabbits is significantly influenced by wild population structure, and thus the timing of the most recent common ancestor (TMRCA) does not correlate with the commencement of the domestication process. Our results demonstrate the significant limitations of using genomic data alone to reconstruct the evolutionary history of a domestic species.

Expanded summary*: Demographic modelling, using the joint site frequency spectrum, is an increasingly popular technique to estimate ancestral population sizes, migration rates and split times. Recent studies have applied these techniques to a diverse range of species and evolutionary contexts—from the out of Africa expansion, to the domestication of rice. Validation of these demographic models, and the software which implement them, is often based on simulated data, or by comparison to estimates made by other techniques. As such, it is unclear how reliable are the inferences drawn from such modelling when applied to real genomic data.

By examining this exemplar dataset, we demonstrate the limitations of current population genetic methods for reconstructing the known demographic history of a species undergoing domestication. This has significant implications for demographic modelling of other domesticated species—particularly where the commencement of domestication is much older, the demographic history is less well defined, or where the progenitor populations are now extinct or restricted to a tiny fraction of their former geographic range (e.g. dogs).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

POB-348

Y-chromosome haplotype diversity in a global sample of free-ranging and companion dogs

Zach T Lounsberry [1], Razib Khan [1], Dan Fulop [1], Ryan H Boyko [1], Spencer Wells [1], Aaron J Sams [2,*], Adam R Boyko [2 3]

[1]Embark Veterinary, Austin, TX, [2]Embark Veterinary, [3]Biomedical Sciences, Cornell University, Ithaca, NY, United States

Abstract: Utilizing 569 Y-chromosome single-nucleotide polymorphisms (SNPs) in over male 2,700 dogs, we have generated the largest known reconstruction of the paternal history of modern dogs. These data are inclusive of village dogs, purebred, and mixed breed dogs. We report 152 total Y chromosomal haplotypes, with novel haplotypes discovered in 9 breeds. Greater numbers of Y-chromosome markers and purebred dogs allow us to resolve the paternal history of unrelated breeds that were previously thought to share Y-chromosome haplotypes.

In addition to increased resolution of paternal phylogeographic history among domesticated dog breeds, the presence of exotic canid haplotypes on the Y chromosome in comparison to autosomal and mitochondrial DNA estimates may indicate historically male-biased gene flow. We report 3 novel Y chromosomal haplotypes in dogs exhibiting signs of recent admixture with North American gray wolves. We detected these haplotypes in Arctic-breed dogs both with and without a strong signal of recent North American gray wolf autosomal ancestry. Wolf-like Y-chromosome types in dogs with little-to-no detectable gray wolf autosomal ancestry indicates admixture with North American gray wolves many generations in the past. Outside of North America, all Basenji Y-chromosome haplotypes surveyed did not cluster with any other domestic dog Y chromosome haplogroup. Rather, they formed a branch sister to the rest of the domestic dog tree. This is possible evidence of male-mediated gene flow from exotic canids, likely Middle Eastern and North African wolves, into modern Basenjis. Interestingly, we detect Y chromosomal gene flow from exotic canids on the edge of the range of domestic dogs, in lineages with deep regional roots. The recent colonization of dogs from Europe to other regions of the world may mask the extent of admixture in the past with the decline of non-European lineages.

The expanded Y-chromosome phylogeny we report allows us to decouple breeds that were previously suspected to represent a single male lineage. This, along with detection of introgression of wild canid Y chromosomes, brings us an improved understanding of the role of sex-biased gene flow and reproductive skew in the patterning of genetic diversity in modern dogs.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

POB-342

Increased genetic variation from ancient polyploidy was preferentially selected during the domestication of Brassica rapa crops

Xinshuai Qi [1,*], Hong An [2], Tara Hall [1], Chenlu Di [1], Eric Lyons [3], J. Chris Pires [2], Michael Barker [1]

[1]Department of Ecology & Evolutionary Biology, University Of Arizona, Tucson, [2]Division of Biological Sciences, University of Missouri, Columbia, [3]School of Plant Sciences, University Of Arizona, Tucson, United States

Abstract: Many crops are polyploid or have a polyploid ancestry. Recent analyses have found that polyploidy often preceded the domestication of crop plants. This observation suggests that increased genetic diversity following polyploidy may have been important during the strong artificial selection that occurs during domestication. Despite the long interest in the connection between domestication and polyploidy, this hypothesis has not been formally tested. *Brassica rapa* crops are renowned for their outstanding morphological diversity that includes oil, root, seed, and leaf crops domesticated during the past 5,000 years. Like all “diploid” vascular plants, *B. rapa* is a diploidized polyploid and experienced many rounds of genome duplication. The most recent genome duplication was a hexaploidization that occurred ~15 MYA. Here, we analyzed transcriptome data of more than hundred representative *B. rapa* cultivated accessions. Using a combination of approaches, we identified more than 3,000 candidate genes associated with the domestication of four major *B. rapa* crops. The candidate gene lists were significantly enriched with genes derived from the ancient hexaploidization event. Further, we found that genes derived from this paleopolyploidy contained ~4X the genetic diversity of the non-polyploid derived genes. These results suggest that genetic variation from ancient polyploidy may have contributed to the diversity of domesticated *B. rapa* crops. Given the distribution of polyploidy throughout the history of flowering plants, our results suggest that the genetic legacy of these ancient whole genome duplications may significantly contribute to adaptation even millions of years later.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

OTH-EG6

Evolutionary genomics of European maize and its American counterparts

Jean-Tristan Brandenburg [1], Tristan Mary-Huard [1], Stéphane Nicolas [1], Alain Charcosset [1], Maud Tenaillon[*]

[1]Génétique Quantitative et Evolution - Le Moulon, Institut National de la Recherche agronomique, Université Paris-Sud, Centre National de la Recherche Scientifique, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France

Abstract: Humans have guided the adaptation of crops to a vast range of climatic and ecological conditions. This is particularly true of maize, which was domesticated in a restricted area of Mexico but now displays one of the broadest cultivated ranges worldwide. Here, we sequenced 67 genomes with an average sequencing depth of 18x to document routes of introduction, admixture and selective history of European maize and its American counterparts. Among our lines, we discovered ove 22 million SNPs. We developed a segmentation method to identify segments of unexpectedly high rate of heterozygosity that point to genes potentially involved in inbreeding depression. Genetic structuring and inferences of historical splits revealed two independent European introductions, with modest bottleneck signatures. Our results further revealed admixtures between distinct sources that have contributed to the establishment of three groups at intermediate latitudes in North America and Europe. We combined differentiation- and diversity-based statistics to identify both genes and gene networks displaying strong signals of selection. These include genes/gene networks involved in flowering time, drought and cold tolerance, plant defense, and starch properties.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

POB-426

DELETERIOUS VARIANTS IN ASIAN RICE AND THE POTENTIAL COST OF DOMESTICATION

Qingpo Liu*, Yongfeng Zhou, Peter L. Morrell, Brandon S. Gaut

Poster: Many SNPs are predicted to encode deleterious amino acid variants. These slightly deleterious mutations may provide unique insights into population history, the dynamics of selection, and the genetic bases of phenotypes. This may be especially true for domesticated species, where a history of bottlenecks and selection may enrich the frequency of deleterious variants and signal a 'cost of domestication'. Here we investigated the numbers and frequencies of deleterious variants in Asian rice (*O. sativa*), focusing on two varieties (*japonica* and *indica*) that may have been domesticated independently and their wild relative (*O. rufipogon*). We investigated three potential signals of a potential cost of domestication in Asian rice: an increase in frequency of deleterious SNPs (dSNPs), an enrichment of dSNPs compared to synonymous SNPs (sSNPs), and an increased number of deleterious variants. We found evidence for all three signals of cost, such that domesticated individuals contained ~3-4% more deleterious alleles than wild individuals. These dSNPs were enriched within low recombination regions of the genome and experienced frequency increases similar to synonymous SNPs within regions of putative selective sweeps. A characteristic feature of rice domestication was a shift in mating system from outcrossing to predominantly selfing. Forward simulations suggest that this shift in mating system may have been the dominant factor in shaping both deleterious and neutral diversity in rice.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

OTH-EG1

Contrasting footprints of selection in purebred and free-breeding dogs

Adam Boyko Boyko [1,*], Jess Hayward [2], Michelle White [2], Jennifer Yordy [2], Ryan Boyko [3], Aaron J. Sams [3]

[1]Cornell University, Ithacha, [2]Cornell University, Ithaca, [3]Embark Veterinary, Austin TX, United States

Abstract: Dogs were the first domesticated animal and have a global population of one billion individuals, including over 500 breeds. While all extant dogs are descended from a single domestication event and thus share certain domestication sweeps, natural and artificial selection have been quite different in free-breeding village dogs, landraces, and modern purebred lines. The phenotypic diversity of purebred lines facilitates mapping of causal variants underlying adaptation, while the large number of free-breeding dog populations offers both an opportunity of local adaptation and an insight into the fitness consequences of causal variants in natural populations. Genomic analysis of diverse dogs shows that even within a species, the genetic architecture of complex traits can vary dramatically across populations depending on the nature of selection in the population. Finally, genomic analysis of mixes of dog breeds and introgression with wolves provides additional insights into the genetic basis of domestication and the costs of closed versus open breeding populations

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

POB-345

Seeking the genetic etiology of domestic behavior in house cats and tame foxes as a window into personality disorders

Carlos Driscoll [1,*], Dave Roberson [2], Clay Stephens [3], Stephen Lindell [1], Anna Kukekova [4], Meredith Yeager [5], Christina Barr [1]

[1]Laboratory of Comparative Behavioral Genomics, NIH/NIAAA/LCBG, Rockville, [2]Cancer Genomics Research Laboratory,, NIH/NCI/DCEG, Bethesda, [3]Genomics GPS, LLC, Guilford , [4]Department of Animal Sciences, The University of Illinois at Urbana-Champaign, Urbana, [5]Cancer Genomics Research Laboratory, NIH/NCI/DCEG, Bethesda, United States

Abstract: Aggression can facilitate access to resources and mating opportunities across mammals, including domestic species, and appears to be under selection. However, in humans, exaggerated aggression is a hallmark of some psychiatric and personality disorders. One approach to determining genetic factors contributing to variation in aggression is the study of the genetics of domestication. At its most basic, domestication is a suite of heritable traits affecting behavior. Among most domestic animal species, there is decreased aggression and the ability to coexist with humans. The domesticated fox (Vulpes vulpes) and the house cat (Felis silvestris catus) may be good candidates for modeling how genetic variation contributes to aggressive behavior in humans. The house cat, relative to its wild ancestor, manifests a suite of heritable behaviors characteristic of domesticates, and the tame Russian silver fox is recognized as a superior model of domestication. However, few studies have attempted to identify intra- or inter-specific variation among these species. Here, we explore a draft whole genome sequence of a domestic cat (DC), a wild Asian leopard cat (ALC) (Prionailurus bengalensis) and interspecies hybrid offspring (ALC X DC) differing in their levels of tameness. We also explore genetic variation in two lines of fox selected for marked differences in reactivity and temperament (tame vs. aggressive).
One domestic cat (DC), one Asian leopard cat (ALC) and interspecies hybrid offspring (ALC X DC) were WGS at approximately 15X. Exploiting the phylogenetically close relationship between the domestic dog and fox, we used a dog – based exon assay (Agilent) to characterize genome-wide protein-coding variation in tame-selected, aggressive selected, and unselected foxes farmed at the Institute of Cytology and Genetics, Novosibersk, Russia and wild-caught foxes from Maryland.
In ALC–DC comparisons, SNPs, some of which are predicted to be potentially deleterious by in silico analysis, were found in the transcribed region of 1400 genes and in the coding region of 158 of those. Dog-on-fox exon pulldown resulted in ~80% on-target capture with ~70% of the targets at >20X coverage, successfully resolving >90% of the sequences expected from a dog-on-dog assay. After filtering, ~500,000 SNPs were called in fox as compared to the Broad dog assembly. Filtering dog vs. fox differences, ~50,000 SNPs were novel in fox, as compared to the 2.5 X 10 6 SNPs reported for dog in the Broad database.
Domestic animals are generally less aggressive than their non-domesticated ancestors. Systems permissive of domestication and underlying a “tame” phenotype range from those involving fear and impulse control to those driving reward and sociality. Between tame and aggressive animals, we identified damaging SNPs in gene systems influencing anxiety-like behavior, transcription control, DNA repair, epigenetic processes, synaptic plasticity/transmission, reward, and circadian rhythms. As these systems can contribute to vulnerability to, or resilience to, human psychiatric disorders, identification of genetic variation among domesticated animals with exaggerated differences in their degrees of tameness may inform us of the human condition and aid in identifying appropriate models for examining treatment response to compounds being developed for the treatment of various psychiatric disorders.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

OTH-EG4

Selective sweep analysis using village dogs highlights the pivotal role of the neural crest in dog domestication

Amanda Pendleton [1,*], Feichen Shen [1], Angela Taravella [1], Sarah Emery [1], Krishna Veeramah [2], Adam Boyko [3], Jeffrey Kidd [1]

[1]University of Michigan, Ann Arbor, [2]Stony Brook University, Stony Brook, [3]Cornell University, Ithaca, United States

Abstract: Dogs (*Canis lupus familiaris*) were domesticated from gray wolves between 20-40kya in Eurasia, yet details surrounding the process and impacts of domestication remain unclear. The vast array of phenotypes exhibited by dogs mirror numerous other domesticated animals species, a phenomenon known as the Domestication Syndrome (DS). Here, we sought to detect signatures persisting in the dog genome following the intensive selective pressures of domestication. Through whole-genome SNP analysis of 43 globally distributed semi-feral village dogs and 10 wolves, we distinguished sweeps associated with domestication rather than modern breed formation. We utilized an F_{ST} selection scan to identify selective sweeps that significantly deviate between dogs and their wild ancestors. In total, we have identified 37 F_{ST} candidate domestication regions (CDRs) that encompass 17.5Mb of the genome and contain 172 genes. SNPs conferring missense mutations were not identified within these CDRs, indicating that gene loss did not have a significant role in the domestication of dogs. Further analysis of these CDRs revealed a significant enrichment of genes linked to neural crest cell (NCC) migration, differentiation and development that can explain many of the DS phenotypes shared across species including decreased jaw size, hairlessness, floppy ears, tameness, and smaller brain size. The involvement of NCCs (transient stem cells that are essential in early embryonic development) would account for the far-reaching and superficially unassociated phenotypic changes in DS. Enhanced starch metabolism through Amylase 2B (*AMY2B*) CN expansion has been previously linked to dog domestication. Closer examination of the *AMY2B* locus revealed complex dog-specific structural variants and an extended selective sweep that highlights a novel candidate in dog domestication. Through droplet digital PCR and read-depth based copy-number estimations, we have validated the presence of both tandem duplications and large-scale (1.9 and 2.0Mb) duplications spanning *AMY2B*. Though no extreme *AMY2B* copy-number expansion was observed in ancient dogs (5,000-7,000 years old), the ancient samples share modern dog haplotypes at the locus and only one of three genomes contained a large-scale duplication. Additionally, we identified an extended (>4.5 Mb) selective sweep that contains *AMY2B* but peaks above the adjacent gene *RNPC3.* Other genes encoding minor spliceosome proteins, like *RNPC3*, have been linked to growth disorders as well as disruption in stem and neural crest cell differentiation. Altogether, these patterns suggest that multiple, complex selection pressures may have been acting on this hallmark domestication locus and that *AMY2B* tandem expansion did not likely occur as a result of initial domestication. Rather, primary selection may have been directed at *RNPC3*, supported by the absence of extreme *AMY2B* copy-number expansion in ancient dogs. We propose that during early dog domestication, selective pressures acting on key genes in the activation and development of the neural crest resulted in aberrant differentiation and localization of NCCs in their final tissues (e.g. ears, jaw, brain), giving rise to many dog phenotypes we observe today. Additionally, genes involved in digestion (such as *AMY2B*) may have been secondarily selected for through hitchhiking effects based on proximity to genes involved in NC development.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

POB-344

Ancient European dog genomes reveal continuity since the Early Neolithic

Laura Botigue, Shiya Song, Amelie Scheu, Shyamalika Gopalan, Amanda Pendleton, Matthew Oetjens, Angela Taravella, Timo Seregély, Andrea Zeeb-Lanz, Rose-Marie Arbogast, Dean Bobo, Kevin Daly, Martina Unterländer, Joachim Burger, Jeffrey Kidd, Krishna Veeramah*

Abstract: Europe has played a major role in dog evolution, harbouring the oldest uncontested Palaeolithic remains and having been the centre of modern dog breed creation. We sequenced the whole genomes of an Early and End Neolithic dog from Germany, including a sample associated with one of Europe's earliest farming communities. Both dogs demonstrate continuity with each other and predominantly share ancestry with modern European dogs, contradicting a previously suggested Late Neolithic population replacement. Furthermore, we find no genetic evidence to support the recent hypothesis proposing dual origins of dog domestication. However, our End Neolithic sample possesses additional ancestry found predominantly in modern Indian, Central Asian and Middle Eastern dogs, which we speculate may be derived from dogs that accompanied humans from the Eastern European steppe migrating into Central Europe. By calibrating the mutation rate using our oldest dog, we narrow the timing of dog domestication to 20,000-40,000 years ago. Interestingly, we do not observe the extreme copy number expansion of the AMY2B gene that is characteristic of modern dogs and has previously been proposed as an adaptation to a starch-rich diet driven by the widespread adoption of agriculture in the Neolithic.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

OTH-EG2

Annuals, Perennials and the Cost of Domestication

Brandon Gaut [1,*]

[1]Ecology and Evolutionary Biology, UC Irvine, Irvine, United States

Abstract: Among thousands of domesticated plants, a major distinction is the difference between annual and perennial life cycles. The domestication of perennials is expected to follow different processes than annuals, with distinct genetic outcomes. Here I will discuss those differences and focus on the fate of slightly deleterious mutations through genetic bottlenecks, with examples from the annual Asian rice (*Oryza sativa*) and the perennial grapes (*Vitis vinifera*). For the former, we show that domestication has increased both the frequency of deleterious mutations that survive the domestication process and the net number of deleterious mutations per individual. These deleterious mutations are enriched within low recombination regions of the genome and within regions of putative selective sweeps. These effects are comparatively less evident in the perennial, largely owing to contrasting population dynamics between the species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

POB-341

Genomic and phenotypic evidence for independent domestication of three grain amaranth species

Karl Schmid [1], Markus Stetter [1,*]

[1]Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany

Abstract: Grain amaranth was an important crop in pre-Colombian agriculture and is currently enjoying a renewed interest because of the favorable nutritional qualities of its grains. Three species of grain amaranths are cultivated in Central (Amaranthus cruentus and A. hypochondriacus) and South America (A. caudatus). Morphological and marker analyses suggested a complex domestication history. A Genotyping-by-Sequencing (GBS) analysis of grain amaranths and putative ancestors and other wild relatives provides was consistent with a model of three independent domestications from different subpopulation of the same ancestral species in different geographic regions of South and Central America. In contrast to classical models of domestication, cultivated A. caudatus showed a higher genetic diversity and a weak domestication syndrome compared to their wild relatives. We hypothesize that domestication is not complete possibly because of because of genetic constraints or recurrent gene flow from sympatric wild ancestors. To further investigate this hypothesis, we created an annotated reference of the 500 Mb A. caudatus genom and resequenced more than 120 individuals that represent a balanced sample of all three grain amaranths and the wild relatives A. hybridus and A. quitensis. A demographic analysis confirm the independent domestication of the three grain amaranths from the same ancestor. To test whether the same genomic regions were affected by domestication we compare genomic regions identified by tests of selection with the location of QTLs identified in a segregating population of A. caudatus and its ancestor A. quitensis. We will discuss our results in the light of current models of plant domestication.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of domestication

POA-420

ENVIRONMENTAL ASSOCIATION IN BARLEY LANDRACES: IDENTIFYING THE GENETIC BASIS OF COLD TOLERANCE

Li Lei [1], Ana Poets [1], Chaochih Liu[*], Paul Hoffman [1], Skylar Wyant [1], Corey Carter [1], Richard Trantow [1], Gary Muehlbauer [1], Fumi Katagiri [1], Peter Morrell [1]

[1]University Of Minnesota, Saint Paul, United States

Poster: Adaptation to environmental conditions, including cold, is extremely important in plants. Barley is cultivated across a broad latitudinal range, from the equator to the Arctic Circle (0-66˚N). This broad range of cultivation is especially noteworthy because the wild progenitor, Hordeum vulgare ssp. spontaneum occupies a narrow range (30-40˚N) at <1500m elevation. The genetic basis of cold temperature tolerance has not been extensively explored, but is amenable to environmental association in barley landraces (primitive cultivars). We report an environmental association study involving 803 landraces genotyped with a 9K iSelect SNP chip. Our results identify at least one strong candidate for freezing tolerance in landraces; the associated SNP is found in the wild progenitor at relatively high elevations. An improved understanding of the genetic basis of adaptation to environmental conditions since the origin of cultivation could provide insight into the nature of adaptation to climate.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-177

Expression of tandem gene duplicates is often greater than twofold

David Loehlin*, Sean Carroll [1]

[1]Howard Hughes Medical Institute and University of Wisconsin-Madison, Madison, United States

Abstract: Tandem gene duplication is an important mutational process in evolutionary adaptation and human disease. Hypothetically, two tandem gene copies should produce twice the output of a single gene, but this expectation has not been rigorously investigated. Here, we show that tandem duplication often results in more than double the gene activity. A naturally occurring tandem duplication of the *Alcohol dehydrogenase* (*Adh*) gene exhibits 2.6-fold greater expression than the single-copy gene in transgenic *Drosophila*. This tandem duplication also exhibits greater activity than two copies of the gene in trans, demonstrating that it is the tandem arrangement and not copy number that is the cause of overactivity. We also show that tandem duplication of an unrelated synthetic reporter gene is overactive (2.3- to 5.1-fold) at all sites in the genome that we tested, suggesting that overactivity could be a general property of tandem gene duplicates. Overactivity occurs at the level of RNA transcription, and therefore tandem duplicate overactivity appears to be a previously unidentified form of position effect. The increment of surplus gene expression observed is comparable to many regulatory mutations fixed in nature and, if typical of other genomes, should shape the fate of tandem duplicates in evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-179

Accurate Transposable Element Patterns and Accumulation in Callorhinchus milii

Laura Blanco-Berdugo [1,*], Nelson Platt [1], David Ray [1]

[1]Biological Science, Texas Tech University, Lubbock, United States

Abstract: Abstract

Transposable Elements (TEs) are DNA sequences that have the ability to replicate and insert themselves in different locations within the host genome. TE insertions are generally deleterious or neutral, however they have also been associated with regulatory mechanisms in the genome. Therefore, identifying their accumulation patterns in vertebrates allows us to better understand the factors associated with genome organization and function. Our analysis of *Callorhinchus milii* (The Australian ghostshark)includes an improved method of TE annotation, showing that at least 34.4% of the genome is composed of TEs. This is lower than prior estimates (40%) of TEs using homology methods. Furthermore, we found that the majority of TEs represented were Non-LTR elements (29%) such as SINEs and LINEs. DNA transposons were found in the genome, though in smaller numbers (2%). We also found that LINE2 and SINE2 were key elements in the TE landscape in the Australian ghostshark genome. Our analysis indicates that the amount of transposable element in the Australian ghostshark genome may have been exaggerated, leading to the inaccurate annotation of the TEs in the genome.

Expanded summary*: Transposable elements (TEs)are DNA sequences that have the abilityto mobilize themselves around the host genome, occasionally havening deleterious effects in the genome. Vertebrates are a very diverse taxon that can be found in different biotopes, from freezing oceans to desserts. Many different lineages have developed specific adaptive characteristics, such as gills, fur, placenta and other to survive in these different environments. It is important to ask, what drove the development of all those traits? The answer is a complex combination of interactive factors driving evolution that has played an essential role in the history of vertebrates (~ 500 million year). Transposable elements have been considered as one of the factors that has help evolution to occur, TEs influence their hosts evolution in different ways such as (i) gene function alteration via insertion (ii) via chromosomal rearrangements and (iii) insertion on genetic material that allows the emergence of genetic novelty (new genes and regulatory sequences). Fishes are the most diverse group of living vertebrates, the Australian ghostshark (*Callorhinchus milii*)

Transposable elements occupy large portions of eukaryotic genomes (de koning, et al. 2011) today we know that they constitute more than half of the DNA in may higher eukaryotes (Fedoroff, N. et al. 2012) and have been previously linked to changes in organismal biology and evolution (Kazazian 2004), it has also been previously proposed that TEs play an important role in lineage specific diversification. The Australian Ghostshark genome was sequenced in 2007 (Venkatesh et al) and proposed as an important model due to the fact that is the smallest genome among all cartilaginous fish. Cartilaginous fish are the oldest living jawed vertebrate, the*Callorhinchus milii* represents a useful reference genome that will contribute to the understanding of the origin and evolution of vertebrates. As such, a through annotation of TEs in this genome is important to fully understand the landscape of the genome and its structure.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG6

Structural variation of grape domestication

Yongfeng Zhou [1,*], Brandon Gaut [1]

[1]Department of Ecology and Evolutionary Biology, University of California, Irvine, Irvine, United States

Abstract: Evolutionary genomics of plant domestication had been widely studied for annual plants, but nominally for perennial plants. Meanwhile, Structural variation(SV) is more and more recognized as an important hidden layer of genomic diversity which has been almost wholly ignored in genomic studies of plant domestication. Most crop adaptation might have been driven by SVs and, as a consequence, current studies of crop adaptation may have missed most of the important genetic variants. The grapevine(*Vitis vinifera* spp. *vinifera*), the most important fruit crop in the world, has been used as a source of wine and food for thousands of years. However, the domestication of grapes haven't been studied using whole genome resequencing dataset. We resequenced samples (Illumina for all and Pacbio for few of them) of grape accessions with its putative wild progenitor (*Vitis vinifera* spp. *sylvestris*) and detected SVs including duplications, deletions, insertions, inversions and translocations based on Illumina sequence signals of read-depth, read-pair and split-read while using Pacbio sequence as a verification. We found important SVs for grape domestication and SVs underlying traits such as fruit color. Interestingly, grape domestication had occurred with a shift of sexual system from dioecy to self-pollination. We found major SVs on chromosome 2 associated with sex determination. Most of SVs at genic regions led to gene structure changes, frame-shift, gene losses and mobile elements insertions and were predicted to be deleterious.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG12

Population reference graphs of complex regions

Jacob Malte Jensen [1,*], Isaac Turner [2], Gil McVean [2]

[1]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark, [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

Abstract: Many genomic regions are too variable for variation to be reliably identified from mapping of short reads to a single reference. This raises a challenge in identification of potential causative variants in association studies and the study of evolutionary mechanisms underlying such variation. A possible solution to this challenge is to create population reference graphs (PRGs) that contain multiple reference sequences. PRGs tie together assembly sequences and variant sequences in a graph representation. Current implementations of PRGs in complex regions such as the MHC rely on multiple sequence alignment in blocks of sequence between and within specified regions such as genes. While these approaches are suitable for identifying small variants they limit detection of structural variation.

We propose a PRG construction algorithm that is not dependent on multiple sequence alignment and provide a data structure that can serve as a generative model for haplotypes by stochastically selecting paths through a graph built from phased haplotypes. This reference representation improves identification of structural variants and inference of new haplotypes. We apply the PRG algorithm to a set of complete MHC haplotypes and show that the complexity grows sub-exponentially with the number of haplotypes, making it feasible to include tens of haplotypes in a single reference. The algorithm uses a heuristic approach to minimize redundancy in the graph and a genome reference sequence can be added to allow lift over of genome annotations. We show that the complexity of the graph grows approximately linearly as a function of variant discovery showing that from a set of core reference sequences addition of variant haplotypes mainly adds variant paths to the graph. The algorithm utilizes the McCortex software distribution to build graphs and describe paths, which enables the user to map directly to the graph and perform variant calling.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-183

Birth of Four Chimeric Plastid Gene Clusters in Japanese Umbrella Pine

Chih-Yao Hsu [1,*], Chung-Shien Wu [1], Shu-Miaw Chaw [1]

[1]Academia Sinica, Biodiversity Research Center, Taipei, Taiwan

Abstract: Many genes in the plastid genomes (plastomes) of plants are organized as gene clusters, in which genes are co-transcribed, resembling bacterial operons. These plastid operons are highly conserved, even among conifers,whose plastomes are highly rearranged relative toother seed plants. We have determined the complete plastome sequence of *Sciadopitys verticillata* (Japanese umbrella pine), the sole member of Sciadopityaceae. The *Sciadopitys* plastome is characterized by extensive inversions, pseudogenization of four tRNA genes after tandem duplications, and a unique pair of 370-bp inverted repeats involved in the formation of isomeric plastomes.We showed that plastomic inversions in *Sciadopitys* have led to shuffling of the remote conserved operons, resulting in the birth of four chimeric gene clusters. Our data also demonstrated that the relocated genes can be co-transcribed in these chimeric gene clusters. The plastome of *Sciadopitys* advances our current understanding of how the conifer plastomes have evolved toward increased diversity and complexity.

Expanded summary*: Before this study, the 25 published cupressophyte plastomes available on GenBank (January, 2016) represent four of the five cupressophyte families but no complete plastome is available for Sciadopityaceae. As a part of our continuing efforts to decipher the diversity and evolution of conifer plastomes, we have completed and elucidated the plastome sequence of *Sciadopitys*. We found that the plastome of *Sciadopitys* is characterized by several unusual features, including shuffling of the conserved plastid operons and re-organization of plastid genes into new chimeric gene clusters. The chimeric gene clusters of *Sciadopitys* provide two novel insights into the evolution of plastomes. First, if the promoter sequences of gene clusters have not been altered after inversions take place, genes of different origins are able to be co-transcribed in the chimeric gene cluster. Second, isomeric plastomes may transcribe and translate thier own proteins to compensate for some deleterious inversions. All these characteristics highlight the fact that the evolution of plastomes may be more complicated than previously thought. The highly rearranged plastome of *Scidaopitys* advances our understanding of the dynamics, complexity, and evolution of plastomes in conifers.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-181

Plastid genome evolution across the rosids

Deise Josely Pereira Goncalves [1,*], Edgardo Manuel Ortiz [1], Beryl Simpson [1], Robert Jansen [1]

[1]Integrative Biology, The University of Texas at Austin, Austin, United States

Abstract: Rosids, containing Vitales, fabids, and malvids, represent about one-third of all flowering plants. Previous studies of plastid genome (plastome) variation in rosids have reported some clades with highly rearranged plastomes (e.g., Geraniaceae and Fabaceae) and others with highly conserved genomes (e.g., Melastomataceae). However, many rosid families still lack any information about plastome organization. In addition, while phylogenetic relationships within rosids have been extensively investigated, the placement of some orders is still debated. We expanded plastome sequencing of rosids by sampling 29 species from 19 families, especially within Myrtales and Zygophyllales. These data were used to compare genome size, GC content, gene content, and the percentage of coding sequences. We also compared gene order using the *Nicotiana tabacum* plastome as a reference because it represents the ancestral genome organization for angiosperms. Among the species investigated, Zygophyllales had the smallest and the largest plastomes. Zygophyllaceae have a reduction or loss of *ycf*1 and *ycf*2 and have the smallest plastomes. Krameriaceae have the largest genomes due to the expansion of the large single copy region caused by repetitive sequences in the intergenic spacers. All Vochysiaceae genera investigated have the ancestral angiosperm plastome organization except *Vochysia* and *Salvertia*, which share an expansion of the inverted repeat that duplicated *ycf*1, *rps*15, *ndh*H and a fragment of *ndh*A. Preliminary phylogenetic analysis based on 58 plastid genes strongly supported Zygophyllales as sister to fabids and Geraniales as sister to malvids; the latter is possibly equivocal and will be further investigated.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-180

Rapid mobile sequence turnover and persistent DNA slippage drive the evolution of mitochondrial genome architecture in yeast

Shujie Xiao [1], Duong Nguyen [1], Baojun Wu [1], Weilong Hao [1,*]

[1]Wayne State University, Detroit, United States

Abstract: Mitochondrial genomes (mitogenomes) are remarkably diverse in genome size and organization, but the origins of dynamic mitogenome architectures are still poorly understood. For instance, the mutational burden hypothesis postulates that the drastic difference between large plant mitogenomes and streamlined animal mitogenomes can be driven by their different mutation rates. However, inconsistent trends between mitogenome sizes and mutation rates have been documented in several lineages. These conflicting results highlight the need of systematic and sophisticated investigations on the evolution and diversity of mitogenome architecture. This study took advantage of the strikingly variable mitogenomes among different yeast species and also within species, examined sequence movement of introns, GC-clusters, tandem repeats and their contribution to mitogenomic variation. Our results show a time-dependent manner of these sequence types in mitogenomic variation, perhaps due to a combination of different sequence turnover rates, variable insertion spaces and selection. Although the pattern of mitogenome expansion and contraction in yeast cannot be explained directly by the mutational burden hypothesis, we observed a positive correlation between mitogenome size and the level of genetic drift, suggesting that mitogenome expansion in yeast is likely driven by different types of sequence movements in a primarily non-adaptive manner.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG4

Understanding the Role of Chromosomal Inversions in Life History Divergence and Local Adaptation in Mimulus

Jenn Coughlan [1,*], John Willis [1]

[1]Biology, Duke University, Durham, United States

Abstract: Despite a growing acknowledgement that chromosomal inversions can play an important role in adaptation, the mechanism by which inversions contribute to adaptation remains unclear. An inversion could suppress recombination between multiple, locally beneficial alleles. This would prevent maladaptive reshuffling of locally adaptive alleles with less-fit, migrant alleles. The recombination-suppression hypothesis has strong theoretical support, but the fact that chromosomal inversions suppress recombination makes empirically testing this hypothesis difficult. We explore a chromosomal inversion which differentiates annual and perennial forms of *Mimulus guttatus*. We survey the LG8 inversion throughout the genus, and find that the annual orientation of the inversion is ancestral, even though perenniality is thought to be the ancestral phenotype. We also find that the inversion is present in the more distantly related perennial species- *M. decorus*, and that it explains similar levels of phenotypic variance for life history traits among species as it does between *M. guttatus* ecotypes, suggesting that the inversion is quite old. Lastly, we explore whether the LG8 inversion captured alleles for perenniality that existed in perennial populations before the inversion evolved- consistent with the recombination suppression hypothesis. We did this by QTL mapping life history traits between annual *M. guttatus* and perennial *M. tilingii* (which lacks the LG8 inversion). We find that the LG8 region is still associated with life history traits in absence of the LG8 inversion- in line with the recombination suppression hypothesis. We argue that a comparative genetics approach can be useful in determining the role of chromosomal inversion in adaptation.

Expanded summary*: A fundamental goal of evolutionary biology is to understand what maintains diversity in nature. Spatial heterogeneity in natural selection can maintain diversity, because different habitats may have different selective optima. Habitat-specific selective pressure can result in local adaptation; the phenomenon by which populations have a fitness advantage at local sites over foreign ones. Curiously, closely related ecotypes often vary in chromosomal arrangement, suggesting that chromosomal rearrangements could play an important role in local adaptation. Studies have shown associations between inversions and a local fitness advantage, but we still know very little about *how* this occurs. An inversion could suppress recombination between multiple, locally beneficial alleles. This would prevent maladaptive reshuffling of locally adaptive alleles with less-fit, migrant alleles in the face of gene flow. While this hypothesis has strong theoretical support, we lack studies testing it in nature. Here we explore the adaptive value of a large chromosomal inversion on LG8 which is associated with life history expression between annuals and perennials in the genus *Mimulus.* This inversion has been previously described to be associated with the majority of genetic variance for life history traits between annuals and perennials of *M. guttatus*, is one of the only genetic markers which differentiate life history variants in natural populations, and is strongly associated with fitness in the field. We first surveyed the presence of the LG8 inversion throughout the *Mimulus guttatus* species complex and beyond and find that the annual orientation of the inversion is likely ancestral, even though annuality is the derived phenotype. We also find that this inversion exists in the more distantly related *M. decorus*- a perennial species which co-occurs with annual *M. guttatus*- and determine that the inversion in this species also controls the majority of life history variance between co-occuring populations of perennial *M. decorus* and annual *M. guttatus*. We interpret these findings to suggest that the LG8 inversion is quite old. Lastly, we explore whether the LG8 inversion has captured multiple alleles for perenniality which pre-date the inversion as a test of the recombination suppression hypothesis. We accomplish this by mapping life history traits between annual *M. guttatus* and the more distantly related *M. tillingii*, which possesses the ancestral phenotype (perenniality) but lacks the LG8 inversion. We find that the LG8 inversion region is still associated with life history traits, despite *M. tilingii* lacking the inversion, in line with the recombination suppression hypothesis. This work will eventually shed light on the mechanistic role of chromosomal inversions in speciation and secondary contact, as well as provide information about the genetic basis of life history variation between annuals and perennials in this group.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-176

Evolutionary forces shaping the distribution of the GSTM1 gene deletion in humans

Marie Saitou [1,*], Omer Gokcumen [1]

[1]Dept. of Biological Sciences, State University of New York at Buffalo, New York, United States

Abstract: The complete deletion of the cellular metabolizing gene, glutathione-s-transferase μ1 (*GSTM1*) is common in human, reaching to major allele frequency in most human populations. However, evolutionary mechanisms through which this deletion has been maintained at high frequencies remain unknown. In this study, we have resolved the haplotype structure of the locus harboring the *GSTM1*-deletion-tag and identified a 9kb haplotype block defined by nine SNVs and the *GSTM1* deletion in East Asian populations. We analyzed this haplotype to investigate the evolutionary neutrality of the *GSTM1* deletion. We found that this haplotype showed unusually elevated frequency in East Asian populations, which is reflected in very high F_{ST} (0.62-0.67) values between YRI and CHB on the *GSTM1* tag region. Indeed, these SNVs show significantly higher F_{ST} as compared to F_{ST} distribution of variants that we randomly selected from chromosome one with matched allele frequencies ($p<0.05$, one-tailed test). To understand the impact of this haplotype to transcriptome, we analyzed gene expression data from Gtex portal. We found that the nine SNVs were significantly associated with differences in expression levels of the *GSTM3* and *GSTM5* genes, which locate downstream of the *GSTM1* gene. Specifically, the East Asian-dominant haplotype decreases expression of the *GSTM5* gene in most tissues, in particular, brain, whereas increases the *GSTM3* expression in muscle ($p<0.01$). This means that the East Asian-dominant haplotype has regulatory effects on the flanking genes. Overall, our data suggest that this haplotype, and consequently the linked *GSTM1* deletion, may have been under non-neutral pressures in East Asian populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-172

Long-read PacBio sequencing reveals the rapid evolution of structurally complex, immune-evasion genes in P. falciparum

Emily Ebel [1,*], Marina McDew-White [2], Tim J.c. Anderson [2], Dmitri Petrov [1]

[1]Biology, Stanford University, Stanford, [2]Texas Biomedical Research Institute, San Antonio, United States

Abstract: Effective vaccines against the malaria parasite, *P. falciparum*, have been hindered by the complexity and volatility of their immune targets. The *P. falciparum* genome encodes over 100 antigenic proteins that are exported to the surface of the red blood cell. Over the course of an infection, single proteins are expressed temporarily and exclusively, creating a perpetual state of "bait and switch" for the host immune system. Furthermore, recent evidence indicates that members of one antigenic gene family, known as *var*, undergo remarkably high rates of structural exchange during asexual reproduction in red blood cells. In this work, we have applied PacBio sequencing to the ancestor and descendants of a 100+ generation mutation accumulation experiment in *P. falciparum*. We have used the resulting long reads (mean length: 7.5 kb) to assemble each *P. falciparum* genome *de novo*, revealing unprecedented structural variation within antigenic gene families. Most of this variation, including dozens of additional copies of annotated genes, has been overlooked by studies that rely on mapping short reads to a reference genome. Between the ancestor and descendants of our experiment, we infer multiple instances of antigenic gene duplication, recombination, and other complex structural changes, which allow us to accurately estimate the rate at which new variation is generated. These results help elucidate this unusual, and clinically problematic, mode of genome evolution in the malaria parasite, as well as demonstrate the utility of new sequencing technologies to approach structural variation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-174

Tectivirus-like Double Jelly Roll Capsids in the Prokaryotic World

Natalya Yutin*, Eugene Koonin

Abstract: The tectivirus-adenovirus lineage unites icosahedral dsDNA viruses with Double Jelly Roll capsid proteins (DJRCP). This lineage is represented by viruses infecting all three cellular domains including bacteria (Tectiviridae), archaea (Sulfolobus turreted icosahedral virus and euryarchaeal proviruses TKV4 and MVV), and eukaryotes (Adenoviridae, Lavidoviridae and the large and giant viruses in the putative order "Megavirales"). We employed a multiple alignment of DJRCP from the Tectiviridae to initiate an exhaustive sequence similarity search in genomic (nr) and metagenomic (WGS) databases. Numerous, highly diverse DJRCP sequences were discovered in various environments and in different genomic contexts. Analysis of the DJRCP-carrying contigs reveals several groups of previously unknown viruses and/or virus-like elements that are unrelated to each other except for the presence of DJRCP and a packaging ATPase.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-218

The age of inversions in Drosophila yakuba and their role in structuring variation

Patrick Reilly [1,*], Julie Peng [1], Peter Andolfatto [2]

[1]Lewis-Sigler Institute for Integrative Genomics, [2]Ecology and Evolutionary Biology, Princeton University, Princeton, United States

Abstract: How do inversions pattern intraspecies genomic patterns of diversity? How do they arise? How old are they? We examine the case of two pairs of overlapping megabase-scale inversions in *Drosophila yakuba*. These inversions cover between 10 and 40% of the chromosome arms, representing a substantial obstacle to recombination, and remain at appreciable frequencies within the species. We identify the breakpoints at the molecular scale by generating and aligning chromosome-scale de novo assemblies of two lines using PacBio sequencing, and examine possible origins of each inversion. Using a variety of published and new *D. yakuba* resequencing datasets, we characterize the population genomic structure introduced by each inversion, and estimate the ages of these inversions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-175

Functional Analyses of Nuclear Encoded Mitochondrial Duplicated Genes with Testis-biased Expression in Drosophila melanogaster

Mohammadmehdi Eslamieh [1,*], Esther Betrán [1]

[1]Biology, UTA, Arligton, United States

Abstract: Most of duplicated nuclear-encoded mitochondrial genes (N-mt genes) in *Drosophila melanogaster* have acquired testis-biased expression. These genes appear to be old, relocated more often than other duplicated genes, have energy-related functions and might have evolved in response to intralocus sexual antagonism. Since males are under intense male-male competition pressures to fertilize females' eggs and do not pass the mitochondria to the offspring, selection might favor mitochondria with high-energy production even though it might also bring more reactive oxygen species (ROS) production. To test this hypothesis, we knock down 39 duplicated genes with testis-biased expression in *D. melanogaster*. Different RNAi lines and different Gal4 drivers were used to study the effects of these genes on viability and fertility. The effects will be compared to CRISPR/Cas9 knockout strains. The knockout protocol has already been setup in the Betrán lab. Also, levels of ROS production will be measured.

Expanded summary*: If the effect of these genes on fertility gets confirmed, they will shed light on human infertility problems.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-170

The birth, expansion, and divergence of tandem transposons in natural populations: A model for the ongoing evolution of complex satellites

Michael McGurk [1,*], Daniel Barbash [1]

[1]Molecular Biology and Genetics, Cornell University, Ithaca, NY, United States

Abstract: Eukaryotic genomes are replete with repeated sequences, dispersed across the genome by transposition (transposable elements, TEs) or in large tandem arrays (satellites). Repeat variation causes reproductive barriers between species, altered transcription, and is implicated in human disease. Yet there is little knowledge of the within-species variation of satellites and the mechanisms by which they arise. Here we analyze Illumina paired-end data sets from the Global Diversity Lines, sequenced genomes from five populations of *Drosophila melanogaster*, and use mixture modelling to identify and infer the structures of tandem repeats. Our analysis successfully captures distinct stages of tandem repeat evolution ongoing within a single species. We find that TEs frequently form low-copy tandems, particularly during periods of active transposition, forming the initial substrate of further expansion. Further, we discovered a rare copy number expansion of a tandem Hobo transposon to ~18 tandem copies, showing that we can also capture the process that yields larger tandem arrays. Finally, we observe substantial copy number variation at established tandem arrays. Analysis of tandem TE junctions suggests that insertion site preference is the mechanism driving tandem formation. We also identify rapid turnover of variants within functional tandem arrays (ribosomal RNA and histone genes), including population specific and emerging variants, but much slower turnover in other tandem repeats. The range of variation we have discovered supports the hypothesis that tandem TEs are an important source of new complex satellite DNAs and demonstrates that questions about population variation in repetitive sequence are tractable using existing population genomic resources.

Expanded summary*: Repetitive DNA is a major feature of eukaryotic genomes, dispersed by transposition events (transposable elements, TEs) or arranged in tandem arrays (satellites). Repeats drive much of the observed genome size variation across species and often occupy structures essential for genome integrity. Satellite variation can create reproductive barriers between closely related species, give rise to selfish processes where centromeric repeats hijack chromosome segregation, and contractions of a human subtelomeric repeat lead to loss of local gene silencing and underlie a major form of muscular dystrophy. Further, variant centromeric repeats have been linked to mitotic segregation errors, which may unmask recessive mutations in tumor suppressor genes, thus representing unexplored risk factors for cancer. Yet there is little knowledge of the within-species variation of satellites and the mechanisms by which they arise.

Unfortunately, the methods used to elucidate the structure and population variation of single-copy sequences often fail when applied to repeated sequences. Consequently, repeats are typically excluded from whole-genome sequencing analyses, and remain severely understudied. Still, each of the tens of thousands of available next-generation sequencing datasets includes millions of repeat-derived reads, and we sought to leverage this abundant information to provide a more complete picture of satellite evolution. We circumvent the bioinformatic challenges repeats pose by first employing an alignment strategy that identifies repeat-derived reads and organizes them in an interpretable way. Second, we recognize that each structure in a sequenced genome produces a distinct distribution of read pairs, so we use mixture modelling to infer the structures (tandems, internal deletions, insertions into unique and repeated sequence) formed by repeats from the aligned data. This provides us with a set of genetic markers in sequences that have up to this point been largely ignored in resequencing projects.

Applying this approach to the Global Diversity Lines, a panel of 85 paired-end sequenced *Drosophila melanogaster* genomes from five populations, we observe evidence of distinct, ongoing stages in the evolution of tandem arrays. We find that TEs frequently form low-copy tandems, particularly during periods of active transposition (such as the recent P-element invasion), forming the initial substrate of further expansion. Further, we capture the process that yields larger tandem arrays, documenting a rare copy number expansion of a tandem Hobo transposon to ~18 tandem copies. Finally, we observe substantial copy number variation at established TE and non-TE tandem arrays. As we often find sequence residing in tandem junctions consistent with known insertion motifs, we suggest that insertion site preference is a major mechanism driving the initial formation of tandem TEs. We also examine sequence variation within tandem arrays. Some repeats (e.g. the centromeric 359-bp repeat) display population structure consistent with the

known demography of the population and a slow rate of variant turnover. However, the ribosomal RNA genes and histone cluster, tandem repeats with known functions, are distinct in that their repeat units display both highly constrained positions as well as positions where variant alleles change in copy number at a rate much faster than observed in other repeats, yielding population specific variants. We are currently engaged in comparing the observed copy number and sequence variation with models of satellite evolution to better tease out the evolutionary forces at work here. Overall, we report substantial population variation at tandem repeats and that TEs provide abundant source material for the emergence of new tandem arrays. Further, by providing a means to identify markers within repeats our method expands considerably the amount of sequence in highly repetitive regions—such as centromeres, piRNA clusters, and degenerate sex chromosomes (e.g. the human Y)—, that can be assayed for population variation using currently available datasets.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-192

Abundant hidden structural variation within and between three Drosophila species revealed by new de novo reference genomes

Mahul Chakraborty [1], Jeffrey Vedanayagam [2], Ching-Ho Chang [3], Colin Meiklejohn [4], Kristi Montooth [4], Kevin Thornton [1], Amanda Larracuente [3], J.J. Emerson [1 5,*]

[1]Ecology and Evolutionary Biology, University of California Irvine, Irvine, [2]Developmental Biology, Sloan Kettering Institute, New York, [3]Biology, University of Rochester, Rochester, [4]Biological Sciences, University of Nebraska-Lincoln, Lincoln, [5]Center for Complex Biological Systems, University of California Irvine, Irvine, United States

Abstract: Mutations that add, subtract, rearrange, or otherwise refashion genome structures often affect phenotypes, though the fragmented nature of most contemporary assemblies obscure them. We solve this problem with high quality, reference-grade genome assemblies for the members of the *Drosophila simulans* clade, the sister group to *Drosophila melanogaster*, which includes *D. simulans*, *D. sechellia*, and *D. mauritiana*. These genomes, assembled from more than 100-fold coverage long reads, are comparable in contiguity and completeness to that of *D. melanogaster*, permitting us to construct an extremely high resolution SV map in all three species. This map reveals abundant variation previously invisible to existing reference genomes and high througput short-read methods, including lineage-specific copy number variants and a nearly complete catalog of transposable elements. On average, each genome possessed more than 400 sequence duplications relative to the *D. melanogaster* genome (428 in *D. simulans*, 420 in *D. mauritiana*, and 436 in *D. sechellia*). Among these duplicates, 20% of the CNVs copied protein-coding genes. We also discover that previous assemblies from the 12 genomes project missed more than two-thirds of the transposable element (TE) content, each genome being composed of ~18% TEs.

Using these resources, we have deciphered ~180Kb of highly repetitive flamenco piRNA cluster, revealing rapid turnover of this structurally complex genomic feature. In *D. simulans* and *D. mauritiana*, we map population samples of short reads to the new reference and reveal abundant polymorphic SV. Finally, we explore structural variation in a sample of five de novo assembled *D. simulans* genomes drawn from the sample above, assembled from 100-fold long reads. All five genomes achieve assemblies comparable in completeness and contiguity to the *D. melanogaster* reference genome. By aligning these genomes to one another, construct a comprehensive map of SVs that reveals a vast amount of hidden variation that exceeds the total variation due to SNPs and small indels. In total, this work represents the highest resolution of comparative population genomics of structural variation in metazoans to date.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-195

Distinct Evolutionary Patterns of Genomic Structural Variation Between Fission and Budding Yeasts

Ahmad Rajeh*, Zhenguo Lin [1]

[1]Biology, Saint Louis University, St. Louis, United States

Abstract: Genomic structural variations are known to contribute to evolution of novel traits and speciation. Comparative studies of structural variations between different organisms could provide new insights into phenotypic diversification and speciation from a genomic perspective. In this study, we employ a strict pipeline to identify and validate deletion, duplication, inversion, and translocation events in 61 populations of the budding yeast *Saccharomyces cerevisiae* and 32 populations of the fission yeast *Schizosaccharomyces pombe*. We observed distinct patterns of genomic structural variations between those two species. Specifically, the genomes of budding yeast populations are much more dynamic than those of fission yeast, in terms of the number of structural variation events per unit of sequence divergence. Second, the distributions of the four types of variants examined are quite different. Furthermore, most of translocation events are mapped to regions proximal to retrostransposon, which are known as high spots of translocations. Since budding yeasts have much more retrotransposons than fission yeasts (50 vs 13), the massive loss of retrotransposons in fission yeast is likely an important factor for the more conserved genome structure among their populations. Our results suggest that unlike nucleotide substitutions, the rates and patterns of genomic structural variation are quite distinct among organisms, and are likely due to their different genome content.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-194

The effect of common inversion polymorphisms In(2L)t and In(3R)Mo on patterns of transcriptional variation in Drosophila melanogaster

Andrew Kern [1,*], Erik Lavington [1]

[1]Genetics, Rutgers University, Piscataway, United States

Abstract: Chromosomal inversions are an abundant class of DNA polymorphism that can have outsized effects on the phenotype of an individual. Inversion variation has been implicated in a broad range of diseases, segregation distortion, local adaptation of natural populations, and speciation. While this is so we have little understanding of the general mechanisms, genic or chromosomal, by which inversions shape phenotypes. Here we use genomic sequence and expression data from the Drosophila Genetic Reference Panel to explore the effects of two cosmopolitan inversions, *In(2L)t* and *In(3R)Mo*, on patterns of transcriptional variation. We demonstrate that each inversion has a significant effect on transcript abundance for hundreds of genes across the genome. Inversion affected loci (IAL) appear both within inversions as well as on unlinked chromosomes. We report that while inversions seem to have more limited regional effects on transcription, there is clear evidence of a break point proximal effect for *In(3R)Mo.* Finally we explore possible functional enrichments among IAL and show that a group of sterol uptake genes seem to be targeted as IAL by a chromatin remodelling protein that is in linkage disequilibrium with an inversion breakpoint. Our results highlight a previously unexplored dimension of chromosomal inversion polymorphism that may lead to profound impacts on fitness.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-188

Genome rearrangement in Euplotes woodruffi: An evolutionary perspective

Yi Feng [1,*], Leslie Beh [2], Laura Landweber [2]

[1]Biological Sciences, [2]Columbia University, New York, United States

Abstract: Genome remodeling during development has influenced vertebrate evolution, for example in shaping the immune system. The aberrant reshuffling of somatic DNA sequences also contributes to diseases like cancer. Ciliates undergo an exaggerated process of programmed genome rearrangement during development via chromosome fragmentation, deletion, and sometimes even programmed inversion or translocation. Previously, our lab has studied genome remodeling in the ciliate *Oxytricha trifallax*. In this study, we used the distantly related spirotrichous ciliate *Euplotes woodruffi* as a new model to study genome rearrangement. First, we sequenced the somatic genome of *Euplotes woodruffi*. Then we used a novel computational pipeline based on split-read mapping to infer the locations of DNA rearrangement junctions in the genome. We specifically focused on inference of the short DNA sequences at these junctions, which play a role in bridging consecutive DNA sequences during genome remodeling. This study will provide us with insight into the evolution of genome architecture and chromosome rearrangements in ciliates.

Expanded summary*: Genome remodeling during development has influenced vertebrate evolution, for example in shaping the immune system. The aberrant reshuffling of genome sequences also contributes to diseases like cancer[1]. Ciliates undergo an exaggerated process of programmed genome rearrangement via chromosome fragmentation, deletion, and sometimes even programmed inversion or translocation.

Ciliates have separate somatic and germline nuclei. In vegetative growth, mainly somatic genome is transcriptionally active. During sexual conjugation, one copy of the zygotic germline micronucleus (MIC) differentiates into somatic macronucleus (MAC)[2], and the genome is rearranged during development. The macronucleus destined sequences (MDS) in MIC will be retained and the internal eliminated sequences(IES) interrupting MDS will be deleted. Previously our lab reported *Oxytricha trifallax* with extremely complicated genome architecture. In *Oxytricha trifallax*, MIC contains diploid long chromosomes and MAC contains about 16,000 gene-sized nanochromosomes[3]. During genome rearrangement, over 90% of the genome is discarded and large scale of unscrambling is needed to form functional genes in MAC[4].

Mechanisms of genome rearrangement has been widely studied. Short identical sequences flanking IES are commonly found in MIC, which are called as "pointers". The same pointer sequences appear at 3' of n_{th} MDS and 5' of $(n+1)_{st}$ MDS in MIC, and only one copy of pointer is retained at that MDS-MDS junction in MAC. Pointers are likely to play a role in rejoining two consecutive MDS sequences but not sufficiently, since they are usually only 2-20bp. The most common 2-nucleotide pointer is TA in *Oxytricha trifallax*, which is the only pointer in *Paramicium* and the most common one among *Euplotes*[5]. Previously the complete investigation of pointers in *Oxytricha trifallax* in our lab showed that scrambled MDS have longer pointers compared to nonscrambled MDS[4]. This suggests the complexity of pointer sequences could reflect the complexity of genome rearrangement.

Genome rearrangement is also regulated by epigenetic mechanisms. In *Oxytricha*, small RNAs mapped to MDS are produced for retention of those sequences, which is opposite to the mechanism using scnRNA which will mark eliminated DNA in *Tetrahymena* and *Paramicium*[6]. Long non-coding RNAs are used as templates for genome remodeling[7]. The different mechanisms in genome rearrangement suggest evolution of genome remodeling.

A 7-bp pointer in *Euplotes octocarinatus* has been reported[8]. From what we have known in *Oxytricha*, longer pointers contain more specific information, which could indicate complicated genome remodeling in *Euplotes octocarinatus*. *Euplotes woodruffi* is closely related to E*uplotes octocarinatus* in phylogenic tree, so we decided to study genome rearrangement using *Euplotes woodruffi* as a new model. Previously our lab investigated the rearrangement process in *Oxytricha trifallax* by aligning MAC genome to MIC genome. However, this classic method requests deep sequencing and good assembly of MIC genome. Here we use a novel computational method only requireing MIC reads which are partially mapped to MAC genome to infer DNA rearrangement junctions. This method will provide an economic way to study genome rearrangement. By comparing features of DNA rearrangement junctions in different ciliates, this study will also provide us with insight into the evolution of genome architecture and chromosome rearrangements.

Bibliography

[1] Campbell, P. J., Yachida, S., Mudie, L. J., Stephens, P. J., Pleasance, E. D., Stebbings, L. A., ... & McBride, D. J. (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*, *467*(7319), 1109-1113.
[2] Prescott, D. M. (1994). The DNA of ciliated protozoa. *Microbiological reviews*, *58*(2), 233-267.
[3] Swart, E. C., Bracht, J. R., Magrini, V., Minx, P., Chen, X., Zhou, Y., ... & Jung, S. (2013). The Oxytricha trifallax macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. *PLoS Biol*, *11*(1), e1001473.
[4] Chen, X., Bracht, J. R., Goldman, A. D., Dolzhenko, E., Clay, D. M., Swart, E. C., ... & Sebra, R. P. (2014). The architecture of a scrambled genome reveals massive levels of genomic rearrangement during development. *Cell*,*158*(5), 1187-1198.
[5] JACOBS, M. E., & KLOBUTCHER, L. A. (1996). The long and the short of developmental DNA deletion in Euplotes crassus. *Journal of Eukaryotic Microbiology*, *43*(6), 442-452.
[6] Yerlici, V. T., & Landweber, L. F. (2014). Programmed genome rearrangements in the Ciliate Oxytricha. *Microbiology spectrum*, *2*(6).
[7] Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T. G., & Landweber, L. F. (2007). RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature*, *450*(7172).
[8] Wang, W., Zhi, H., Chai, B., & Liang, A. (2005). Cloning and sequence analysis of the micronuclear and macronuclear gene encoding Rab protein of Euplotes octocarinatus. *Bioscience, biotechnology, and biochemistry*, *69*(3), 649-652.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-193

How do inversions affect recombination? A study in the gibbon

Debora Y C Brandt [1,*], Lucia Carbone [2], Jeffrey D Wall [3]

[1]University of California, Berkeley, [2]Oregon Health & Science University, Beaverton, [3]University of California, San Francisco, United States

Abstract: Individuals that are heterozygotes for inversions (heterokaryotypes) might suffer detrimental effects due to the generation of meiotic products containing large deletions, or lacking centromeres. This happens in Drosophila through the formation of a loop in the pairing of homologous chromosomes during meiosis. When a single recombination event happens within that loop, inviable gametes are formed. Only double recombination events within the loop generate viable gametes, therefore recombination is lower in inversions, and particularly lower near the inversion breakpoints. However, there is evidence that in some mammals (including humans) heterokaryotypes might not form a loop during meiosis, but instead the homologous chromosomes do not pair along the inversion, impeding recombination completely. Two distinct predictions are thus raised: i) if mammals form a homologous pairing loop during meiosis, then recombination will be lower near the edges of the inversion, and higher towards the center (due to rare double recombinants); ii) if mammals do not form homologous pairing during meiosis, then recombination will be shutdown along all the inversion. Lower recombination rates cause the alternative karyotypes to accumulate differences, thus increasing genetic diversity in the region. We test these alternative hypotheses analyzing genetic diversity along the sequence of a single gibbon (Hylobates) individual sequenced at high coverage, who is heterozygous for a long and old inversion (shared accross species). If gibbons form the meiosis loop, we expect to see higher diversity near the inversion breakpoints than in the center of the inversion. Otherwise, we expect higher diversity along all the inversion.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POB-416

DYNAMIC LANDSCAPE OF HUMAN L1 TRANSPOSITION REVEALED WITH FUNCTIONAL DATA ANALYSIS

Di Chen [1 2,*], Marzia Cremona [3], Zongtai Qi [4], Robi Mitra [4], Francesca Chiaromonte [2 3 5], Kateryna Makova [2 6]

[1]Intercollege Graduate Degree Program in Genetics, The Huck Institutes of the Life Sciences, [2]Center for Medical Genomics, The Huck Institutes of the Life Sciences, [3]Department of Statistics, The Pennsylvania State University, University Park, [4]Department of Genetics and Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, United States, [5]Sant'Anna School of Advanced Studies, Pisa, Italy, [6]Department of Biology, The Pennsylvania State University, University Park, United States

Poster: The Long Interspersed Element-1 (L1) is one of the most active human Transposable Elements (TEs), which constitutes over 17% of the human genome. Characterizing the transpositional activity of L1s and their interactions with the genomic landscape is critical for understanding genome evolution and function. However, to date, the dynamics of L1 integration and fixation has not been studied comprehensively. We performed a genome-wide investigation of L1 transposition using three large datasets of *de novo*, polymorphic, and (human-specific) fixed L1s (15,996, 956, and 995 elements, respectively). We measured 50 genomic features (e.g. nucleotide composition, histone modifications, gene expression level, etc.) at high resolution in the flanking regions of these elements and in control regions (L1-free). We next used Interval-Wise Testing (IWT), a novel Functional Data Analysis tool, to contrast their landscapes at multiple scales and identify signatures of L1 integration and fixation. Our results suggested that *de novo*, polymorphic, and fixed L1s in the human genome are characterized by unique genomic landscapes, with different features acting at specific locations and scales. While *de novo* L1 integrations tend to occur in regions with high G/C content, open chromatin and elevated gene expression levels, fixed L1s tend to concentrate in regions with low exon content and transcriptional repression. The genomic landscape of polymorphic L1s is somewhat similar to that of fixed L1s, though the signals from IWT are less significant. These findings confirm that the genome-wide distribution of L1 elements is far from random, and delineate the role of genomic features in both insertion and fixation preferences. As the next step in our analysis, we are using multiple Functional Logistic Regression to quantify the joint effects of the genomic features selected by IWT. In a summary, here for the first time, we investigate L1 transposition on the genome-wide scale and in the evolutionary framework, while considering interactions with an extensive range of genomic features. This study sheds light on the dynamics of TE landscape and will advance our understanding of the structure, evolution and function of the human genome.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-424

STATISTICAL METHODS FOR DETECTING CRYPTIC COPY NUMBER VARIATION FROM POPULATION-LEVEL NEXT GENERATION SEQUENCING DATA

Tyler Linderoth [1,*], Rasmus Nielsen [1]

[1]University Of California Berkeley, Berkeley, United States

Poster: We have developed a likelihood method for detecting paralogy using next generation sequencing data from multiple individuals based on a likelihood ratio (LR) test of whether a particular genomic site has been duplicated. For ease of analyzing long, contiguous lengths of sequence we use these LRs as emissions in a Hidden Markov Model in order to obtain discrete coordinates of duplication. Through simulation, we demonstrate that this method is powerful at detecting paralogy for even small sample sizes with low sequencing coverage (e.g. 10 individuals at 2X average coverage is usually more than sufficient) where other methods for detecting paralogy break down. We applied these methods to the 1000 Genomes low-coverage human dataset to reveal yet unrecognized duplicated regions in the human genome in addition to *de novo* assemblies from non-model organisms. The ability to detect paralogy in this manner has broad application ranging from studying the evolutionary significance of duplication to its potential to confound population genetic inference when paralogous regions are misassembled.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POB-413

GENOMIC ANALYSIS OF TRANSPOSABLE ELEMENT AMPLIFICATION IN WOODPECKERS AND ALLIES (AVES: PICIFORMES)

Stephane Boissinot*, Jospeh Manthey [1]

[1]New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Poster: Birds' genomes typically have a lower diversity and number of transposable elements (TEs) in their genomes than other vertebrates. Most bird species' genomes have less than 12% TE content, while the downy woodpecker (*Dryobates pubescens*) genome consists of > 20% TE content. We sequenced genomes for representatives of all lineages within the Piciformes, including jacamars, barbets, toucans, honeyguides, and woodpeckers and asked the following questions: (1) When did TEs increase in abundance within Piciformes genomes? (2) Is the pattern of amplification similar across woodpeckers? (3) How many groups of TEs expanded? We found that several families of the CR1 non-LTR retrotransposon simultaneously amplified in the genomes of Piciformes, totaling about 20% of their genomes. We also discovered that the amplification of CR1 in two independent lineages, the woodpeckers and the toucans + barbets, occurred independently. In both cases, TE accumulation appears to have preceded bursts in the diversification rate of the hosts (woodpeckers and new world barbets + toucans). These results contrast with general patterns of genome evolution in birds, where TEs are generally found in small abundance across the genome because of a reduced activity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POB-411

OHANA: DETECTING SELECTION IN MULTIPLE POPULATIONS BY MODELLING ANCESTRAL ADMIXTURE COMPONENTS

Jade Cheng [1 2,*], Rasmus Nielsen [1 2]

[1]Departments of Integrative Biology and Statistics, University Of California, Berkeley, Berkeley, United States, [2]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

Poster: One of the most powerful and commonly used methods for detecting local adaptation in the genome is the identification of extreme allele frequency differences between populations. In this paper we present a new maximum likelihood method for finding regions under positive selection. The method is based on a Gaussian approximation to allele frequency changes, which allows it to account for interbreeding between populations and retains high power when the test populations are admixed. It can also simultaneously and efficiently compare multiple populations. We evaluate the method using simulated data and compare it to methods using summary statistics. We also apply it to a human genomic data set and identify loci with extreme genetic differentiation between major geographic groups. Most of the genes identified are previously known selected loci relating to hair pigmentation and morphology, skin and eye pigmentation, including the top two genes: EDAR and SLC25A5. In contrast to previous genomic scans, we include data from Aboriginal Australians, which provide us with additional power to detect selection specific to East Asians. Using this method, we can identify new candidate loci - like CASC15, involved in melanoma suppression - and narrow down on the likely causal SNPs in previously reported candidate regions - like KCNB2, involved in various neurological functions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-184

New methods for detecting selection on structural variants and introgressions

Luca Ferretti [1,*], Alex Klassmann [2], Carla Giner-Delgado [3], Sara Guirao-Rico [4], Thomas Wiehe [2], Mario Cáceres [3], Mathieu Joron [5], Guillaume Achaz [6], Sebastián Ramos-Onsins [4]

[1]The Pirbright Institute, Woking, United Kingdom, [2]University of Cologne, Cologne, Germany, [3]Universitat Autonoma de Barcelona, Bellaterra, [4]CRAG, Barcelona, Spain, [5]CEFE, Montpellier, [6]UPMC & Collège de France, Paris, France

Abstract: Both structural variants (SVs) and introgressions are important sources of genomic and phenotypic variability in populations, often with strong phenotypic effects. For this reason, many polymorphic SVs in natural population could be under selective pressure. However, detecting ongoing selection on SVs is a challenging task due to their impact on local patterns of variation and recombination.
Here we present a new set of methods to detect violations of neutrality in several simple SVs, including inversions, deletions and insertions, as well as introgressions from close species. These methods are based on the frequency spectrum of linked sites - similarly to Tajima's D and other classical tests - and can detect both positive and balancing selection on SVs. We present some applications of these methods to SVs in human populations and to introgressions in Heliconius butterflies, which could be easily extended to other organisms.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-182

The expected length of identical by descent blocks on the genome over generations in a finite panmictic population.

Mathieu Tiret*

Abstract: With the highly dense genomic data available nowadays, because of New Generation Sequencing techniques, ignoring linkage between genes would result in a huge loss of information. Therefore, developing multi-locus measures have become substantial in genome scan analyses. In this article, we have focused on the identity by descent (IBD) shared by genomic tracts, in a Wright-Fisher population assuming panmixia, no selfing and drift as the only evolutionary pressure. In such populations, crossovers break apart the ancestor haplotypes as the time runs and form smaller blocks at each generation. Some blocks, and eventually all of them, become identical by descent (hence called IBD blocks). Using computational simulations and the theory of junctions, we provide here a prediction of the mean length of IBD blocks, with which it is now possible to accurately predict this length at any generation after founding.

Expanded summary*: With the highly dense genomic data available nowadays, because of New Generation Sequencing techniques, ignoring linkage between genes would result in a huge loss of information. Therefore, developing multi-locus measures have become substantial in genome scan analyses. In this article, we have focused on the identity by descent (IBD) shared by genomic tracts, in a Wright-Fisher population assuming panmixia, no selfing and drift as the only evolutionary pressure. In such populations, crossovers break apart the ancestor haplotypes as the time runs and form smaller blocks at each generation. Some blocks, and eventually all of them, become identical by descent (hence called IBD blocks). Using computational simulations and the theory of junctions, we provide here a prediction of the mean length of IBD blocks, with which it is now possible to accurately predict this length at any generation after founding.

IBD block length is a measure of the extent of what is transmitted altogether from one generation to another. Having an accurate prediction of this measure is the first step for incorporating neighbouring in population genetics: what does it mean to be close on the genetic map? Is there any absolute definition, or do we have to nuance depending on the case? For instance, in LD pruning, we arbitrarily state that two locus are close if they are of a certain threshold distance. The theoretical framework developed here aims to nuance that threshold relatively to the population genetic context (demography, genetic map specificity, ...) of the case and provide a population-dependent threshold rather than an arbitrary one.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG10

Multiple Independent Retroelement Insertions in the Promoter of a Stress Response Gene

Miriam Merenciano [1,*], Vivien Horváth [1], Josefa González [1]

[1]Institute of Evolutionary Biology (CSIC-Pompeu Fabra University), Barcelona, Spain

Abstract: Transposable elements (TEs) can produce a broad range of structural variants including the generation of new transcripts by providing new transcript start sites or affecting the polyadenylation signal choice, for example. In a previous work, we found nine independent TE insertions in the promoter region of a stress-response gene in *Drosophila melanogaster*. These nine TEs are clustered in a small 368bp region and they all belong to the *roo* family. Moreover, one of these insertions affects the transcription start site of its nearby gene and it is associated with increased tolerance to cold-stress. In this work, we have screened 15 natural populations and we have found 11 new independent *roo* elements inserted in the same promoter region. We also found that some of these insertions are associated with small duplications. In order to know whether all these insertions are functionally relevant, we analysed their sequences and, besides transcription factor binding sites, all of them also provide a polymerase pausing sequence. We are currently checking whether the *roo* insertions could be causing polymerase pausing as this is one of the mechanisms that allow genes to rapidly respond to stress. Finally, we have also tested whether other *roo* elements inserted in the reference genome also cluster in the promoter region of genes. We did not find any other cluster suggesting that cluster formation is not a common feature of *roo* insertions. Our results will allow us to know whether all the insertions in the cluster are functionally relevant and their molecular mechanisms.

Expanded summary*: Understanding the relationship between mutations and its consequent phenotypic effect is needed if we are to understand environmental adaptation. The use of transposable elements (TEs) as tools to analyze adaptive phenotypes is not so well established. Until recently, the mutagenesis ability of TEs has been considered detrimental for genome viability, due to the negative effect of the majority of TE-induced mutations in the host fitness. However, it has been recently shown that TEs are a considerable source of adaptive mutations in *Drosophila melanogaster*. The goal of our lab is to characterize the ecological adaptive effects of putatively adaptive TEs found in *D. melanogaster*. In order to do that, we characterized the phenotypic effect of individual TE insertions and we also analyze the molecular mechanism behind those functional changes.

It is known that TEs can produce different types of structural variants including the generation of new transcripts. In a recent work, we discovered nine naturally occurring independent TE insertions in the promoter region of a cold-stress response gene named *CG18446*. The nine TE insertions are clustered in a small 386 bp region and they all belong to the *roo* family. We thus investigated whether the different insertions were functionally equivalent by performing 5'-RACE, gene expression, and cold-stress survival experiments. We found that different insertions have different molecular and functional consequences. However, only one *roo* insertion named *FBti0019985* provides an alternative transcription start site to *CG18446* and is associated with increased viability in nonstress and in cold-stress conditions. This could be an example of how a TE insertion is affecting the structure of its nearby gene and, it is associated with the organism adaptation to cold conditions.

In this work we used a population genomic approach in order to know how many insertions are located in this insertion cluster. We screened 15 natural populations from three continents and we found 11 new insertions located in the promoter region of *CG18446*. Each individual insertion is present at relatively low population frequencies, ranging from 1% to 18%. However, the majority of strains analyzed contain one of these 20 *roo* insertions suggesting that this region might be evolving under positive selection. Strikingly, in a small number of strains we found duplications in the flanking regions of *roo* insertion fragments. We suggest that this type of structural variants could be associated with the presence of the *roo* elements. Furthermore, we are currently performing run-on assays to know whether these insertions are functionally playing a role in the expression of *CG18446*. Analysing their sequences we found that all of them have a conserved polymerase pausing sequence. It is known that some stress-responsive genes, as well as some developmental genes, can have a paused Polymerase II (Pol II) in their proximal promoter region. This paused Pol II engages in transcription but pauses immediately downstream of the transcription start site. Upon activation by stress, it is able to rapidly transcribe the gene. Thus, our hypothesis is that these insertions are providing the capacity for Pol II to pause and be ready for transcription under stress conditions. This fact can give an adaptive advantage for the flies having one of these insertions because they

could deal with the stress more rapidly. Moreover, we have analyzed by PCR the promoter region of all the genes that have in the reference genome a *roo* insertion inside the 5'UTR, overlapping the 5'UTR or placed not more than 1kb far away from the beginning of the gene. So far, we have not found any evidence of the existence of another insertion cluster. Thus, it suggests that the cluster formation is not a common feature of *roo* elements.

In *D. melanogaster*, repeated insertions of TEs have only been described in the proximal promoters of a particular gene class: *hsp* genes. In this work we have discovered the first *roo* insertion cluster in the proximal promoter region of a non-*hsp* gene. In summary, our results showed that different *roo* insertions in the same gene promoter region might have different structural, molecular, and functional consequences. Thus, the description of complex regions should be followed by functional analysis of the structural variants if we want to elucidate which ones are functionally relevant.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-187

Investigating recombination in the mating type locus of the Chlamydomonas reinhardtii

Jaspreet Duggal [1,*], Rob Ness [1 2]

[1]Biology, University of Toronto, Mississauga, [2]Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada

Abstract: Suppressed recombination in the sex-determining regions of homologous chromosomes allows for the association of sex-specific alleles. Divergence and expansion of the non-recombining sex determination regions has lead to the evolution of sex chromosomes in some species. However, suppressed recombination reduces the ability of natural selection to purge harmful mutations and fix adaptive changes. The model green alga, *Chlamydomonas reinhardtii*, contains a genetic mating-type determination system that shares parallels with early stages in the evolution of sex chromosomes. Recombination in this region was long thought to be suppressed but recent work has shown unexpectedly low divergence between the two mating type alleles and evidence for gene conversion. In this study, we provide a fine scale characterization of the recombination landscape and the consequences of recombination in the *Chlamydomonas reinhardtii* mating-type locus. We have used whole-genome sequence from 24 isolates collected from nearby localities. The sampling scheme used here has facilitated an in-depth analysis of recombination rate variation and its consequences across the entire MT-locus. Using strict parameters, we have identified homologous regions between the MT+ and MT- and have measured rates of recombination along these regions. We will present data that examines how variation in the rate of recombination in this region has altered patterns of molecular evolution including GC content and linked selection.

Statement: I am in my last year of completing my undergraduate degree in Biology at the University of Toronto Mississauga. Over the four years, the most exciting experience has been to be take part in research. My courses in molecular and evolutionary biology introduced me to the field of evolutionary genomics which is the focus of my thesis. Recent work in this area has inspired me to pursue research at a graduate level. I believe the SMBE conference will expose me to new ideas and experiences and enhance my research background.

The most appealing feature of the conference is the mentorship opportunity it presents for an undergraduate who hopes to pursue future research in the field of molecular biology. As a female researcher from an under-represented minority background, I feel this opportunity as well as the experience of presenting my work at an international scientific conference will provide a strong foundation for my research career. My interactions with established researchers will be valuable in my path to learn more about the research and the scientific process. Although I have just begun my research career, I believe my work on which addresses the consequences of recombination between mating types can make valuable contributions to the area of evolutionary genomics and the SMBE conference in 2017.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-190

Chromosomal translocations during the evolution of the Puerto Rican parrot (Amazona vittata)

Juan C. Martinez-Cruzado [1,*], Edwin G. Ramírez-Aponte [1], Isaac Benmaman-Santiago [1], Alexandra Calderón-Landrón [1], Paola Correa-Alfonzo [1], Thays Fontánez-Collazo [1], Carlos Rivera-López [1], Jenniffer Quiñones-Colón [1], Olga Vera-Colón [1], Ingrid T. Rivera-Pagán [1], Taras K. Oleksyk [1]

[1]Biology, University of Puerto Rico at Mayagüez, Mayagüez, Puerto Rico

Abstract: The Puerto Rican parrot (*Amazona vittata*) is the only native parrot species in Puerto Rico and the last one remaining in any of the U.S. territories. It was listed as an endangered species in 1967, and by 1975 its population had diminished to 13 individuals. Its genome was sequenced (coverage = 94.6X; N50 = 2,056,511 pb) through a community-funded project with the aim of contributing to the U.S. Fish and Wildlife Service-led recovery program. We are using a variety of methods to reconstruct the *A. vittata* genome architecture. BLAT alignments of two-to-three 25 kb sequence segments of the 3,096 largest parrot scaffolds to the chicken genome were performed, identifying 51 instances of sequences within a scaffold matching different chicken chromosomes with scores higher than 1000, and thus suggesting evolutionary translocations. Whereas these BLAT alignments suggested 36 translocations between two of the largest 9 autosomes or chromosome Z, Fluorescent In Situ Hybridizations (FISH) of chicken paints for these chromosomes revealed only three chr. 6-7 and one chr. 8-9 translocations. Primers designed to perform PCRs encompassing the putative translocation joints confirmed two chr. 6-7 and one chr. 8-9 translocations. They also showed the presence of a chr. 3-9 translocation specific for the genus *Amazona* that was not found by FISH, and a chr. 2-15 translocation specific for Psittacidae. We are now developing an alternative strategy, which consists on characterizing scaffold G-C content. Because an inverse correlation exists between chromosome size and G-C content, the size of the chromosome where putative translocations between a macro- and a micro-chromosome are located can be predicted, and later confirmed by studying the genomes of other Amazon or parrot species. The predictions must be normalized for G-C content variation along chromosomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-185

A new method for structural variation discovery using PacBio long reads

Sarah Kingan [1,*], Aaron Wenger [2]

[1]Apps Lab, [2]PacBio, Menlo Park, United States

Abstract: Recent studies in humans have demonstrated that PacBio long-read, SMRT sequencing has dramatically improved sensitivity for structural variants (≥ 50 bp) compared to short-read DNA sequencing. SMRT sequencing detects around 20,000 structural variants in a human genome compared to 4,000 detected with Illumina sequencing, primarily due to the improved ability of PacBio long reads to span repeat regions and large insertions. These early results suggest that there is widespread polymorphic structural variation in the human population as has been observed with single nucleotide variants. Here we apply a new method for SV detection using low-coverage (~10x) SMRT sequencing in multiple strains of the fruitfly *D. yakuba*. As in humans, *D. yakuba* has abundant structural variation. The Cameroon strain has 2,061 deletions and 2,030 insertions relative to the reference genome with an enrichment of insertions on the X chromosome relative to the autosomes and a dearth of both events in coding sequence. Approximately 10% of variants involve the Helitron repeats DNARep1_Dyak (282 deletions, 225 insertions) and DNARep1_DM (121 deletions, 167 insertions), demonstrating widespread repeat activity in the *D. yakuba* population. Only 115 insertions contain sequence homology to flanking sequence, indicating that the majority of inserted sequences are not simple tandem duplications. PacBio long reads are particularly sensitive to moderately-sized insertions of non-repetitive sequence, which are poorly detected by short-read and hybridization-based screens. This new approach enables the characterization of polymorphic structural variation in a cost effective manner.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-191

Genomes as documents of evolutionary history: a probabilistic macrosynteny model for the reconstruction of ancestral genomes

Yoichiro Nakatani*, Aoife McLysaght [1]

[1]Trinity College Dublin, the University of Dublin, Dublin, Ireland

Abstract: The human genome harbors ~7000 ohnolog genes, or duplicates that derive from whole-genome duplication (WGD) events at the origin of vertebrates. They are often associated with human diseases, and it is therefore important to make a comprehensive catalog of ohnologs. High-confidence identification of ohnologs hinges on synteny analysis and inference of pre- and post-WGD ancestral genomes, but the ancient timing of teleost and vertebrate WGD events impedes high-accuracy reconstruction of ancestral gene order.

Instead of focusing on gene order conservation, we developed a probabilistic model of macrosynteny (i.e., conserved linkage or chromosome-scale distribution of orthologs), devised variational Bayes algorithms for inferring the structure of pre-WGD genomes, and studied estimation accuracy by simulation. Our high-resolution reconstruction revealed previously overlooked small-scale rearrangements, necessitating a revision to previous views on genome evolution in teleost and vertebrates. Specifically, it has been argued that teleost lineages had remarkably low rates of structural change for a long evolutionary time after the teleost WGD, while several early vertebrate lineages underwent massive structural changes in a short evolutionary time. Our reconstruction refines this view by showing that (1) some chromosomes accumulated small-scale inter-chromosomal rearrangements even in slowly evolving teleost genomes, and (2) by contrast, some chromosomes might have experienced exceptionally strong structural constraints, preserving ancestral vertebrate linkage even in rapidly changing genomes.

Expanded summary*: The human genome harbors ~7000 ohnolog genes, or duplicates that derive from whole-genome duplication (WGD) events at the origin of vertebrates. While most redundant copies have been deleted shortly after WGD, human ohnologs have been retained for more than 500 million years due to their functional constraints: Specifically, it has been shown that retained ohnologs are associated with dosage-balanced genes [1], for which changes in gene dosage have deleterious effects. Indeed, human ohnologs are less likely to be affected by copy-number variation [1], and their copy-number change is frequently associated with disease. For example, in Down's syndrome, ~75% of previously reported candidate genes are dosage-balanced ohnologs [1], and in cancer, ~70% of amplified and overexpressed cancer genes [2] are ohnologs. These examples show that improved annotation of ohnologs contributes to a deeper understanding of disease causing genes.

Construction of a comprehensive catalog of ohnologs hinges on synteny analysis and inference of pre- and post-WGD ancestral genomes. Several formal algorithms have been developed for reconstructing gene order in pre-WGD ancestral genomes, and applied to WGD events in yeasts and plants. However, to the best of our knowledge, those algorithms that target WGDs in yeasts and plants have never been successfully applied to teleost and vertebrate genomes, presumably due to high rates of gene loss and genome rearrangement.

Instead of reconstructing ancestral gene order, we developed a probabilistic model of macrosynteny (i.e., conserved linkage or chromosome-scale distribution of orthologs), devised variational Bayes algorithms for inferring the structure of pre-WGD genomes. Our high-resolution reconstruction revealed previously overlooked small-scale rearrangements, necessitating a revision to previous views on genome evolution in teleost and vertebrates [3, 4] as described in Abstract.

In this study, we have reconstructed the structure of a pre-WGD genome by employing a variational Bayes approach that was originally developed for inferring topics from millions of text documents [5]. Interestingly, comparison of the macrosynteny and topic model algorithms suggests that macrosynteny can be regarded as documents on ancestral genome structure. From this perspective, the present study would seem to provide a textbook example of the prevalent metaphor that genomes are documents of evolutionary history.

References:
1. T. Makino, A. McLysaght, PNAS. 107, 9270–9274 (2010).
2. T. Santarius, et al., Nat Rev Cancer. 10, 59–64 (2010).
3. O. Jaillon et al., Nature. 431, 946–957 (2004).
4. Y. Nakatani, et al., Genome Res. 17, 1254–1265 (2007).
5. D. M. Blei, Commun. ACM. 55, 77–84 (2012).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-178

Repetitive DNA recapitulates evolutionary history in Drosophila

Ian Caldas [1,*], Kevin Wei [2 3], Jullien Flynn [2], Sarah Lower [2], Daniel Barbash [2], Andrew Clark [1 2]

[1]Department of Biological Statistics and Computational Biology, [2]Department of Molecular Biology and Genetics, Cornell University, Ithaca, [3]Department of Integrative Biology, University of California at Berkeley, Berkeley, United States

Abstract: Satellite DNA is composed of long, fast evolving, highly repetitive genomic regions that are important for proper functioning of cellular processes, such as chromosome segregation during mitosis and meiosis. However, due to their repetitive nature, such regions have proven challenging to assemble and annotate from typical short-read sequencing datasets. Here we used a software package, kseek, to *de novo* identify and quantify tandemly repeating units using unassembled (raw) Illumina short-read sequencing data for 9 *Drosophila* species. Differential PCR amplification of fragments produces GC-content bias in read abundance, masking the presence of repeats at the extremes of GC content. We demonstrate improved ability to detect such repeats by using PCR-free library preparation techniques and by implementing a variant of the Benjamini-Speed approach to correct for GC-bias. The correction resolves most differences in repeat count estimation among different library preparations of *D. melanogaster*. Our repeat data were then used to ask whether the satellite sequences alone carried an accurate phylogenetic signal for the genus by applying both parsimony and distance methods. Our tree topologies match the widely accepted consensus for the genus (Nearest-Neighbor Interchange distance between true and satellite-based trees = 1). Our results pave the way for more computational approaches to analyzing difficult-to-annotate repeated DNA and take us a step closer to explicit modeling of satellite evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG13

Rainbow trout genome assembly reveals a double inversion harboring a complex multigenic switch for alternative life-history phenotypes

Devon E Pearse [1], Nicola Barson [2,*], Torfinn Nome [2], Guangtu Gao [3], Matthew A Campbell [1], Alicia Abadía-Cardoso [1], Eric C Anderson [1], David E. Rundio [1], Thomas H. Williams [1], Kerry A. Naish [4], Matthew Baranski [5], Thomas Moen [6], Sixin Liu [3], Matthew Kent [2], David R. Minkley [7], Marine S. O. Brieuc [4], Simen Rød Sandve [2], Michael R. Miller [8], Kobi Baruch [9], Alvaro G. Hernandez [10], Gil Ben-Zvi [11], Doron Shem-Tov [11], Omer Barad [9], Ben Koop [7], John Carlos Garza [1], Steven T. Lindley [1], Gary H. Thorgaard [12], Yniv Palti [3], Sigbjørn Lien [2]

[1]Fisheries Ecology Division, Southwest Fisheries Science Center, , National Marine Fisheries Service,, Santa Cruz, United States, [2]Centre for Integrative Genetics, Norwegian University of Life Sciences, Aas, Norway, [3]National Center for Cool and Cold Water Aquaculture, USDA/ARS, Kearneysville, [4]School of Aquatic and Fishery Sciences, University of Washington, Seattle, United States, [5]Marine Harvest, Bergen, [6]AquaGen, Trondheim, Norway, [7]Department of Biology, University of Victoria, Victoria, Canada, [8]Department of Animal Science, University of California, Davis, United States, [9]NRGENE LTD CITY, Ness-Ziona , Israel, [10]Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, Urbana, United States, [11]Nrgene ltd city, Ness-Ziona , Israel, [12]School of Biological Sciences and Center for Reproductive Biology, Washington State University, Pullman, United Kingdom

Abstract: The importance of chromosomal inversions in adaptation is becoming increasingly evident. Using a de novo chromosome anchored genome sequence and whole genome resequencing, we characterised a large, near chromosome wide (~80Mb), double inversion in rainbow trout. We report that the inversion complex shows a latitudinal frequency cline, and in the southern portion of the range plays a role in determining whether individual trout migrate to sea, or mature directly in freshwater. The centromere is contained within one of the inversions, the flanking regions of which are enriched for genes containing fixed missense mutations and contain multiple candidate genes related to the photic control of circadian rhythm, steroidogenesis and adiposity. The decision to migrate or remain resident is highly sex-biased, with females gaining an advantage from migrating while males benefit from remaining resident. The effect of the inversions on migratory tendency shows sex-dependent dominance, a mechanism by which sexual conflict can be reduced in the absence of morphologically differentiated sex chromosomes. This sexually antagonistic selection provides a mechanism for the maintenance of the inversion polymorphism.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG8

The effect of spatially varying selection on transposable element insertions in Drosophila

Jeffrey Adrion*, David Begun [1], Matthew Hahn [2]

[1]Department of Evolution and Ecology, University of California Davis, Davis, [2]Department of Biology, Indiana University, Bloomington, United States

Abstract: Natural populations often exist in spatially diverse environments and may experience spatial variation in the strength and targets of natural selection over their ranges. This spatially varying selection can shape both the evolution of genome architecture and adaptation to the environment. Drosophila provides an excellent opportunity to study the effects of spatially varying selection in natural populations, as both *D. melanogaster* and *D. simulans* have recently (within the last 500 years) been introduced in North and South America, having since colonized the bulk of both continents. Previous studies have identified candidate single nucleotide polymorphisms (SNPs) that are potential targets of spatially varying selection, and have described broad patterns of SNP variation along clines in North America and Australia. Here, we investigated how spatially varying selection impacts another important source of genomic variation, transposable elements (TEs). TE insertion dynamics shape genome evolution and, in multiple instances, have been shown to be the causative mechanism underlying adaptation to the environment. Here, we discover clinal TE insertions in whole-genome pooled-population sequence data sampled from eight populations of *D. melanogaster* and nine populations of *D. simulans* from the Americas. We characterize parallel variation in the distribution and allele frequency of individual TEs and TE families between species and we explore associations between TEs and features of the host genome. Our results shed light on how TE insertions shape genome evolution in Drosophila and potentially facilitate environmental adaptation.

Expanded summary*: A central aim in evolutionary biology is to elucidate the genetic bases for local adaptation. One fruitful approach has been to sample individuals along geographic transects—such as latitude, longitude, or altitude—that vary predictably in abiotic (e.g., temperature, precipitation, ultraviolet radiation) and biotic (e.g., species biodiversity, levels of competition) conditions. Evaluation of variation along such transects enables the identification of clines, broadly defined as a predictable geographic gradient in a measurable genotypic (e.g., allozyme or allele frequencies) or phenotypic (e.g., body size, thermal tolerance) character. Sampling clines provides unique benefits, and can potentially attenuate some of the confounding effects of demography, which may be difficult to control for when sampling populations from patchy landscapes. For example, gene flow should be more predictable along clines, thus making it easier to identify adaptive from non-adaptive differentiation. Clines are often predictable and replicable to a degree that variation sampled from patchy landscapes is not: for example, a cline along a coastal latitudinal transect can potentially be replicated on multiple continents. Such patterns of differentiation repeated among clines provide evidence of parallel adaptation. Finally, properties of a cline—such as the width, slope, and shape—can also inform inferences about underlying demographic and selective forces. Although adaptation to spatially varying selection has been evaluated for decades using phenotypic data and genetic data from a small number of candidate loci, the recent abundance of whole genome data provides an opportunity to discover novel causative genetic variants—beyond those previously identified by candidate gene studies.

The genus *Drosophila* provides an excellent opportunity to study the effects of spatially varying selection in natural populations, as both *D. melanogaster* and *D. simulans* have recently (within the last 500 years) been introduced in North and South America, having since colonized the bulk of both continents. *D. melanogaster* has traditionally been the subject of abundant studies on clinal variation in phenotypic traits, inversion polymorphisms, and single loci. Moreover, recent studies utilizing whole-genome sequencing have identified candidate single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) that are potential targets of clinal selection, and have described broad patterns of clinal SNP and CNV variation in both North America and Australia.

The movement of transposable elements (TEs) represents another important source of genetic variation: TEs have been implicated in resistance to viral infection and insecticides in *D. melanogaster* and resistance to insecticides in the mosquito *Culex pipiens*. Moreover, comparing variation between closely related sympatric species is a valuable method to detect the effects of natural selection, as homologous traits may display parallel responses to similar underlying selection pressures. Investigations of clinal patterns in TE variation have, thus far, been limited to exploring only a few TE families within *D. melanogaster*. Our investigations are the first to characterize the nature of genome-wide TE insertions in two species of Drosophilid flies sampled along parallel

geographic transects. In this research we describe broad patterns of TE variation shaped by spatially varying selection, we characterize parallel patterns of TE variation between species, and we seek to identify candidate TEs underlying local adaptation. Our results shed light on how TE insertions shape genome evolution in *Drosophila* and potentially facilitate environmental adaptation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-202

Rapid regulatory divergence and dosage compensation in a non-recombining avian autosome with incipient degeneration

Dan Sun [1,*], Iksoo Huh [1], Donna Maney [2], Soojin Yi [1]

[1]School of Biological Sciences, Georgia Institute of Technology, [2]Department of Psychology, Emory University, Atlanta, GA, United States

Abstract: The white-throated sparrow (*Zonotrichia albicollis*), a common winter bird across North America, provides an invaluable resource to study phenotypic dimorphism. In this species, two plumage morphs – tan- and white-striped (hereafter called tan and white) – coexist in the population but differ dramatically in their behaviors. Specifically, the white birds display significantly more territorial aggression and less parental care than their tan conspecifics during the breeding season. Previous karyotype and fluorescence *in situ* hybridization studies have revealed a large rearrangement capturing ~1,000 genes on the second chromosome, causing the observed phenotypic divergence; white birds are heterozygous for the inversion (ZAL2/ZAL2^m, in which ZAL2 represents the non-inverted second chromosome, and ZAL2^m represents the inverted version) and tan birds are homozygous (ZAL2/ZAL2). Our genome sequencing of a rare super-white bird homozygous for the inversion (ZAL2^m/ZAL2^m) shows an excess of nonsynonymous substitutions and of radical amino acid changes, but very few disrupted gene copies on ZAL2^m, suggesting incipient degeneration resulting from recent suppression of recombination. However, transcriptome analysis of multiple tan and white birds reveals a large number of genes differentially expressed between ZAL2 and ZAL2^m. Surprisingly, for some of those genes, we found evidence for dosage rebalance between morphs. These observations demonstrate that the evolution of expression divergence and dosage compensation may predate the degeneration of protein-coding sequences. To our knowledge, our study is the first one to report dosage compensation for a minimally degenerated autosome.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-201

Dynamics of copy number variation in evolving yeast populations

Stephanie Lauer*, Gunjan Gala, David Gresham

Abstract: Copy number variants (CNVs) are a class of large-effect alleles that contribute to adaptive evolution in diverse scenarios. CNVs have played a key role in human evolution and de novo CNVs contribute to a range of human diseases including cancers. In microbial evolution, CNVs are repeatedly observed under conditions of strong selection. Despite the ubiquity of CNVs, fundamental questions concerning the dynamics and diversity of these adaptive alleles remain unsolved. Previous studies have found that CNVs containing genes that encode high-affinity transporters including the general amino acid permease, *GAP1*, are frequently selected in *Saccharomyces cerevisiae* during experimental evolution in chemostats. We have developed a novel GFP-reporter assay that allows sensitive detection of CNVs in mixed populations using flow cytometry. I used this reporter to perform high-resolution, real-time characterization of *GAP1* dynamics in evolving populations and found two temporally distinct phases. Early duplication events are highly reproducible across replicates, but later dynamics are complex and CNVs can become unstable in some populations. To determine the number of independent amplification events that occur, I am combining my reporter with a barcode lineage-tracking method. Using whole-genome sequencing and breakpoint mapping to identify sequence signatures at novel CNV junctions, we aim to define the diversity of CNV alleles in populations under strong selection.

Expanded summary*: Copy number variants (CNVs) comprise duplications or deletions of genomic segments ranging from 50 – 10^6 base pairs. In humans, de novo CNVs introduced each generation are more numerous than point mutations and among individuals, between 4.8-9.5% of the genome contains CNVs. Increases or decreases in gene copy number due to CNVs result in altered mRNA and protein abundance, which can have dramatic effects on cell physiology. Germ-line CNVs underlie a range of human diseases including Crohn's disease, autism and several developmental disorders. Somatic CNVs are frequently found in cancerous cells: nearly 40% of cancer-related genes are found in CNVs. CNVs are a class of large-effect alleles that can drive rapid phenotypic diversification and adaptation in animals, plants, and microbes. Despite the importance of CNVs for phenotypic variation, disease, and evolution, the molecular basis underlying the generation and selection of these alleles is not well understood. The overall goal of my research is to define the role of CNVs in adaptive evolution by determining 1) the dynamics with which they are generated and selected, and 2) how different molecular mechanisms and genomic features contribute to allelic diversity of CNVs.

In microbial evolution, CNVs are repeatedly observed under conditions of strong selection. For example, during nutrient limitation in chemostats, there is frequently selection for CNVs that include the gene responsible for transporting the limited nutrient. This was first observed in experiments with *Escherichia coli* limited for lactose, *Saccharomyces cerevisiae* in phosphate-limited environments and *Salmonella* in different carbon limitations. CNVs containing these genes confer large fitness advantages in specific nutrient-limited environments, but are difficult to characterize and isolate in large, complex evolving populations. To overcome this challenge, I developed a novel fluorescent reporter that allows sensitive detection of CNVs in mixed populations using flow cytometry. This assay allows "visualization" of CNVs with unprecedented temporal resolution, enabling us to monitor the generation and selection of CNVs in real time during adaptive evolution.

I used this reporter to study CNV dynamics for the gene encoding GAP1, the general amino acid permease, in evolving populations of *S. cerevisiae*. GAP1 transports amino acids, including glutamine, and gene duplication occurs when cells are limited for glutamine as the sole nitrogen source. Conversely, when urea is the limiting nitrogen source, *GAP1* expression confers a fitness cost and there is positive selection for loss of *GAP1*. Previous experiments from our laboratory have shown that this deletion is mediated by non-allelic homologous recombination (NAHR) between long terminal repeats (LTRs) flanking *GAP1*. During adaptive evolution in chemostats, cells with a *GAP1* deletion appeared and swept to fixation by generation 125 in one of nine experimental populations. Genome sequencing suggests that this deletion was mediated through NAHR as previously described. I also maintained nine experimental populations under glutamine limitation and detected *GAP1* duplications around generation 70. *GAP1* amplification alleles sweep to fixation in several populations, but CNVs are unstable in other populations. These results suggest that early dynamics are predictable,

based on mutation rate and population size. However, as the mean population fitness increases over time, subpopulations with CNVs may be outcompeted. Genome sequencing confirms *GAP1* duplication, and sequence signatures at breakpoints indicate that several types of amplification alleles (including aneuploidy) occur in different populations. Since chemostat populations are large (10^9-10^{10} cells) and CNV mutation rates are high (between 1×10^{-10} to 3.4×10^{-6} amplifications per cell per division), multiple types of amplification alleles are likely to occur within populations and undergo clonal interference. To determine the number of independent amplification events that occur, I am combining my CNV reporter with a barcode lineage-tracking method.

The strong selective advantage conferred by CNVs in nutrient-limited environments, and the repeatability of these adaptive outcomes, make experimental evolution in chemostats an ideal system for studying CNVs. My work provides new insights into the dynamics and mechanisms underlying this important, but understudied, class of genetic variation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG11

Chromosome rearrangements in a recently emerged facial tumour in Tasmanian devils

Maya Kruger-andrzejewska [1], Janine Deakin [1,*]

[1]Institute for Applied Ecology, University of Canberra, Canberra, Australia

Abstract: Tasmanian devils are currently under the threat of extinction in the wild due to not one but two deadly transmissible facial tumours. Extensive cytogenetic and genomic analyses has been performed on the first transmissible facial tumour (DFT1), which arose over 20 years ago, demonstrating that two chromosomes have experienced extensive rearrangement and most likely started from telomeres on these chromosomes becoming critically short. The second facial tumour (DFT2) was first discovered in 2014 and is karyotypically distinct from DFT1. As the reference genome sequence for the Tasmanian devil consists of over 30,000 unordered scaffolds, we are unable at present to reliably identify chromosome rearrangements using a genomics approach. Instead, we have used a molecular cytogenetics approach to characterise the chromosome rearrangements in four different DFT2 samples. Like DFT1, this second transmissible tumour appears to have arisen from the telomeres becoming critically short, enabling the fusion of two chromosomes. Chromosome 1 is also the most rearranged chromosome in DFT2 as it is in DFT1 but the rearrangements are different between the two tumours. Chromosome 1 has also been highly rearranged in the course of marsupial evolution, making it tempting to suggest that this region of the genome in marsupials is more susceptible to breakage and rearrangement than other chromosomes. Comparisons between DFT2 samples reveal major differences, suggesting tumour evolution is occurring. The study of this more recently emerged transmissible tumour may provide insight into the early stages of the formation of the highly scrambled DFT1 genome, making ti particularly interesting to continue tracking the evolution of chromosome rearrangements in DFT2.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-208

A maximum likelihood approach to estimating the insertion frequencies of transposable elements from population sequencing data

Xiaoqian Jiang [1,*], Haixu Tang [1], Michael Lynch [1]

[1]Indiana University, bloomington, United States

Abstract: Transposable elements (TEs) are capable of expanding to a large fraction of the genomes of many eukaryotic species. As TEs are important for genome evolution, developing appropriate method to estimate TE variations in populations is fundamental to understand evolution. Here, we develop a novel computational method for genotyping the TE insertion polymorphisms in a population of a diploid organisms using paired-end sequencing reads, and a maximum-likelihood (ML) method for estimating the allele frequencies of TE insertions as well as the magnitudes of selection. Simulation data showed that, our method can effectively correct the bias in estimating the allele frequencies of TE insertions if not sufficient coverage of paired-end reads of the sequenced individuals. The method can detect negative selection on TE insertions when the selection strengths are relatively strong. In these cases, false-positive rates in the detected selection on TE insertions are nearly zero, while the false-negative rates may be high (i.e., a substantial fraction of selected TE insertions may be missed by the ML method). Application of our method to the genomic sequencing *Daphnia pulex* population data showed that on one hand, most TE insertions present in the reference *D. pulex* genome have high frequencies in the population and exhibit no selection; on the other hand, among the novel TE insertions, a substantial insertions are demonstrably experiencing purifying selection.

Expanded summary*: Transposable elements (TEs) are a class of DNA components found in most eukaryotic genomes, also play critical roles in shaping the host genomic architecture. TE insertions with highly deleterious effects will be removed by purifying selection, while only TE insertions with sufficiently mild deleterious effects on the host genome have opportunities to spread in a population. The estimate of TE insertion frequency in a population reveals important information about the population demography and the strength of natural selection, and thus is fundamental in population genomics. Although some computational methods have been developed to detect TE insertions, the high sequence similarity among TE elements in the same family may introduce bias in TE detection. Developing appropriate mathematical methods is crucial to address this issue for genome-wide surveys of the frequencies of TE insertions in populations. In this work, we present a maximum-likelihood (ML) method for estimating the frequencies of TE insertions from population genome sequencing data. Combined with a bioinformatics pipeline to detect TE insertions using NGS data, our method can simultaneously estimate the allele frequencies at the polymorphic sites of TE insertions and the bias in TE detection based on the observed genotypes of the sequenced individual genomes. We evaluated our methods on simulated data. The results showed that unbiased estimates of frequencies of TE insertions can still be obtained if relatively low coverage ($\geq$5) of paired-end reads of the sequenced individuals, without requirement of the Hardy-Weinberg equilibrium (HWE). We further showed that our method can identify potential TE insertions under natural selection, and in some cases can predict the strength of the selection correctly. Application of this method to a natural *D. pulex* population reveals that the allele frequency distributions of novel and known TE insertions follow different patterns, and a subset of TE insertions are under negative selection.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-214

Long-term balancing selection on chromosomal variants associated with crypsis in stick insects

Doro Lindtke [1 2,*], Kay Lucek [3], Romain Villoutreix [2], Timothy E. Farkas [4], Aaron A. Comeault [5], Rüdiger Riesch [6], Stuart R. Dennis [7], Zach Gompert [8], Víctor Soria-Carrasco [2], Patrik Nosil [2]

[1]Biological Sciences, University of Calgary, Calgary, Canada, [2]Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom, [3]Botany, University of Basel, Basel, Switzerland, [4]Ecology and Evolutionary Biology, University of Connecticut, Storrs, [5]Biology, University of North Carolina, Chapel Hill, United States, [6]School of Biological Sciences, University of London, Egham, United Kingdom, [7]Aquatic Ecology, EAWAG, Kastanienbaum, Switzerland, [8]Biology, Utah State University, Logan, United States

Abstract: How color polymorphisms are maintained within populations over long periods of time remains debated, as genetic drift and directional or stabilizing selection are expected to eliminate variation. We study the genetic architecture and maintenance of phenotypic morphs that confer crypsis in *Timema* stick insects, using genotyping-by-sequencing data and phenotypic information from 21 populations. Two highly divergent chromosomal variants that span megabases of sequence are present within all populations and are significantly associated with color polymorphism. These variants exhibit highly reduced gene flow and probably diverged hundreds of thousands or more generations ago. We detect heterokaryotype excess and signs of balancing selection acting through the species' history. A third chromosomal variant in the same genomic region is associated with dorsal pattern polymorphism and arose more recently. Our results suggest that large-scale genetic variation associated with crypsis has been maintained for long periods of time by potentially complex processes of balancing selection. In ongoing work we are now investigating the structural architecture of chromosomal variants, and we discuss their importance for adaptation to heterogeneous habitats.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-215

Accurate and fast detection of complex and nested structural variations using long read technologies.

Fritz Sedlazeck*

Abstract: Structural variations (SVs) substantially contribute to phenotypic diversity, evolution and genetic disease (e.g. cancer). Short-reads has proved invaluable to recognizing copy number variations and other simple SVs, although has been limited for detecting most other SVs because of repetitive elements and other limitations of short reads. The advent of long-read technologies, such as PacBio or Nanopore sequencing that now routine produce reads over 10,000bp, offer a more powerful way to detect SVs. However, available methods often lack precision and sensitivity when working with highly erroneous reads, especially for complex or nested SVs.

Here we present Sniffles, a method for detecting all types of SVs from long-read sequencing data, while controlling for false signals from the mapping. A unique feature of Sniffles is detecting nested SVs, such as inversions flanked by deletions, which we now commonly detect in several samples. Furthermore, Sniffles offers read-level phasing to study complex breakpoints. Using real and simulated data, we demonstrate the enhanced ability of Sniffles to detect SVs over existing methods (PBHoney) or short read methods such as Lumpy, Delly, or Manta. We further introduce a new long-read mapping method called NGM-LR to enhance the accuracy of Sniffles and reduce the false discovery of SVs even further.

Working with genuine PacBio and Oxford Nanopore reads with human cancer samples (SKBR3), healthy human samples (GiaB), and other species, we show how Sniffles and NGM-LR reduces the coverage, and therefore cost, required per sample. Sniffles and NGM-LR are available open-source at Github.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-199

Structural variants appear to be transient, and have strong effects on quantitative traits and reproductive isolation in fission yeast.

Daniel Jeffares [1 2,*], Clemency Jolly [3], Mimoza Hoti [2], Doug Speed, Liam Shaw [2], Charalampos Rallis [4], Francois Balloux [2], Christophe Dessimoz [2], Jürg Bähler [2], Fritz Sedlazeck [5]

[1]University of York, York, [2]University College London, [3]The Francis Crick Institute:, [4]University of East London, London, United Kingdom, [5]Baylor College of Medicine, Houston, United States

Abstract: Large structural variations (SVs) in the genome are harder to identify than smaller genetic variants but may substantially contribute to phenotypic diversity and evolution. Here we analyze the effects of SVs on gene expression, quantitative traits, and intrinsic reproductive isolation in the yeast *Schizosaccharomyces pombe*. We establish a high-quality curated catalog of SVs in the genomes of a worldwide library of *S. pombe* strains, including duplications, deletions, inversions and translocations. We show that copy number variants (CNVs) frequently segregate within closely related clonal populations, are weakly linked to single nucleotide polymorphisms (SNPs), and show other genetic signals consistent with rapid turnover. These transient CNVs produce stoichiometric effects on gene expression both within and outside the duplicated regions. CNVs make substantial contributions to quantitative traits such as cell shape, cell growth under diverse conditions, sugar utilization in winemaking, whereas rearrangements are strongly associated with reproductive isolation. Collectively, these findings have broad implications for evolution and for our understanding of quantitative traits including complex human diseases. See: http://biorxiv.org/content/early/2016/10/03/047266

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG2

Ancestral genome reconstruction and ancient polyploidy in vertebrates

Aoife Mclysaght*, Yoichiro Nakatani

Abstract: This is a sketchy abstract because it's only January and I will be presenting this talk in July on work that is currently ongoing, so I can't write a detailed description because I don't have one yet. :-)
We have a new method to reconstruct ancestral genomes and this has allowed us to get more accurate and more complete inference of ohnolog relationships. I will present these results as well as interpretations of the patterns of evolution and retention of ohnologs. It'll be really interesting. I promise.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-196

Pleiotropic noncoding regulatory elements are under purifying natural selection

David W Radke [1 2 3,*], Daniel J Balick [1 2 3], Jae H Sul [4], Sebastian Akle [1 2 5], Matthew Maurano [6], Robert Green [2 3], John Stamatoyannopoulos [7], Shamil Sunyaev [1 2 3]

[1]Department of Biomedical Informatics, Harvard Medical School, [2]Division of Genetics, Brigham and Women's Hospital, Boston, [3]Broad Institute, Cambridge, [4]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, [5]Organismic and Evolutionary Biology, Harvard University, Cambridge, [6]Institute for Systems Genetics, New York University, New York City, [7]Department of Genome Sciences, University of Washington, Seattle, United States

Abstract: Assessing the role of natural selection on genetic variation in noncoding regions has been previously difficult, because most work has only utilized SNP variation, of which there are no clearly defined loss-of-function (LoF) noncoding variants. Genomic deletions, however, provide a powerful LoF noncoding model by removing the nucleotides altogether. Using regulatory annotations from tissues and cell-types characterized as part of the Roadmap Epigenomics Project, along with deletions from large-scale population data, we examine differences in deletion allele frequency as it relates to their regulatory element overlap. Regulatory annotations that share the same genomic position can serve as a proxy for the cellular pleiotropy of that position. We hypothesize that regulatory loci exhibiting highly-pleiotropic effects (i.e., loci with regulation across many diverse tissues) should be under stronger purifying selection than cell-type-specific or nonfunctional loci. We develop a statistical method to account for covariance between the tissues/cell-types, thereby providing a per-base-pair normalized count of co-localized regulatory activity. Analyzing the locus pleiotropy overlapped by each of the noncoding deletions, we find a statistically significant shift in the allele frequency spectrum (AFS) towards rare alleles not only for deletions overlapping regulatory loci versus nonfunctional loci, but also for deletions overlapping highly-pleiotropic loci versus cell-type-specific loci, confirming our hypothesis. We interpret these results as evidence of the widespread action of purifying selection on noncoding regulatory elements, the strength of which is determined by the corresponding amount of pleiotropy. These findings allow more rigorous noncoding functional interpretation for use in medical or experimental studies.

Expanded summary*: Previous work in trying to understand natural selection in noncoding regions has attempted to utilize SNPs altering regulatory binding-site motifs, as well as conserved DNA sequence motifs. Yet motif-disrupting nucleotides are not often easy to identify, as many binding events are non-binary, in that SNP alterations may only disrupt the efficacy of binding, not abolish it altogether. Also, while conservation analysis of DNA sequence can provide a powerful indication of functional regulatory localization, it cannot explain the underlying functional basis of those nucleotides. Our approach utilizes CNV deletion events occurring in noncoding genomic regions that have previously accumulated along the human lineage in both European (EUR) and African (AFR) super-populations, as identified by the 1000 Genomes Project Phase 3 SV subgroup. Additionally, we validate our findings in a European cohort of the Alzheimer's Disease Neuroimaging Initiative (ADNI), creating our own deletion callset from WGS data from 752 individuals. We utilize the inherent heterozygous loss-of-function property of deletion events, and compare allele frequency distributions of deletions overlapping various noncoding regulatory features, looking for shifts in the frequency distribution towards rare alleles, thereby suggesting the prior action of purifying natural selection on specific regulatory functions.

We use regulatory annotations from the NIH Roadmap Epigenomics Mapping Consortium, in particular DNase1 hypersensitivity (open chromatin) and H3K4me1 (enhancer) peaks from 25 primary diverse tissues and cell-types in order to quantify, at base-pair resolution, cellular pleiotropy across the genome. To account for the correlation structure between the tissue annotations, we adapt the PSIC method (Sunyaev, 1999), which was originally designed to identify 'independent counts' of amino acid substitutions from protein alignments across taxa, to instead derive a per-base-pair normalized count of co-localized regulatory activity derived from only

the 25 human tissue alignments. Developing this count into a pleiotropy ratio (ranging from 0 to 1), we identify non-functional (ratio of 0), cell-type-specific (low ratio), and highly pleiotropic (high ratio) regulatory loci across the genome.

Overlaying the many deletion coordinates with the corresponding pleiotropy ratios, we find that deletions that overlap more base-pairs of either DHS or enhancers are more likely to be rare (taking into account genomic covariates such as GC content, recombination rate, repeat proportion, distance to TSS, among others). This helps to functionally explain the classic CNV observation that longer deletions tend to be more rare in populations. Furthermore, we find that deletions that overlap highly pleiotropic loci (versus those that only overlap tissue-specific loci) are even more likely to be rare, suggesting that purifying natural selection is operating to preserve genome regulatory activity on both the "horizontal" (number of base-pairs) and "vertical" (per-base-pair pleiotropy) axes of the noncoding genome.

Our conservation-independent approach allows identification of putative high-function pleiotropic noncoding loci, which helps facilitate research projects aimed at identifying core regulators on a genome-wide scale. With the advancement of technologies like CRISPR/Cas9 that can be readily applied to noncoding function-guided knock-in/out experiments, our work provides high-value targets for genome-wide experimentation for hypothesized loss-of-function loci or adaptive loci. As well, because our method can rapidly assess the potential regulatory impact of any de-novo or rare CNV deletion, advances in pre-natal genetic counseling or childhood diagnosis could be further accelerated. Additionally, as advances in regulatory genomics continue broadly, including for diverse taxa, our method is robust to the addition of more diverse tissues/cell-types and so new data can be readily assimilated into the existing pleiotropy analysis method, thereby providing finer functional resolution of noncoding loci on an on-going basis.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG9

The pervasiveness of gene conversion facilitates interspecies gene flow

Katharine L Korunes [1,*], Mohamed A. F. Noor [1]

[1]Biology Department, Duke University, Durham, NC, United States

Abstract: Interspecies hybridization is common, yet hybridizing species often remain distinct despite opportunities for gene flow. Recombination barriers such as chromosomal inversions can contribute to the maintenance of distinct species. Inversions can hamper gene flow by reducing recombination in inversion heterozygotes, where single-crossover products are prevented within inversions. However, evidence for non-crossover gene conversion within inversions suggests that the degree of recombination-suppression imposed by inversions remains poorly understood. There are extremely limited empirical estimates of gene conversion rates with respect to inversions, so despite the prevailing thinking, inversions may be quite ineffective at preventing gene flow. To test the efficacy of inversions at maintaining linkage-disequilibrium in hybrids, I have quantified gene conversion rates in experimental crosses of the naturally-hybridizing species pair *Drosophila pseudoobscura* and *D. persimilis.* Whole-genome sequencing reveals individual gene conversion tracts, and with these data, I can assess rates of gene flux. I detect gene conversion rates within inverted regions of species hybrids that are at least as high as rates based on within-species LD data. Given such high rates of gene conversion in hybrids but high sequence differentiation between species, other features of this system must maintain the distinction of these species besides the recombination barrier provided by chromosomal inversions.

Expanded summary*: Newly formed species often persist as distinct from their sister taxa despite opportunities for gene flow via hybridization. Current prominent speciation models posit that hybridizing species are more likely to persist when differences in chromosomal arrangement suppress recombination in hybrids. When an individual is heterozygous for a chromosomal inversion, single crossovers within the inversion do not yield viable gametes. However, genetic exchange within inversions can still occur via non-crossover gene conversion. An understanding of the size, distribution, and frequency of gene conversion is crucial to understanding the extent to which inversions hamper gene flow, yet little direct empirical data exist on gene conversion with respect to inversions in species hybrids. Using whole-genome sequence data, I have detected gene conversion throughout the collinear and inverted regions of experimental crosses of *Drosophila pseudoobscura* and *D. persimilis*. This naturally-hybridizing species pair provides a system where I can assess whether the rate of gene conversion differs within vs. outside inversions, among areas within inversions (e.g. near breakpoints), and among inversions of varying divergence (inter-/intra- specific). I detect gene conversion both within and around inversions, and my data indicate that gene conversion rates within inverted regions of hybrids are at least as high as published rates based on within-species population genetic data. Given the high observed rates of gene conversion in hybrids but high sequence differentiation between species, the persistence of these species is not solely attributable to recombination suppression by inversions. Evolutionary forces such as natural selection against gene flow must also be present to maintain these species.
Broad Significance: If gene conversion rates are sufficiently high and vary independently of inversions, then, despite the prevailing thinking, inversions may be quite ineffective at keeping hybridizing species genetically distinct. Theoretical work and limited empirical evidence suggests that gene conversion has the potential to undermine inversions as barriers to gene flux. The present study lends direct, genome-wide empirical evidence to this theory. These data provide the basis for the first direct analysis of gene conversion with respect to inversions in hybrids, and our estimates of gene conversion rates show that gene conversion occurs regularly within inverted regions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-198

The TeddyPi pipeline: Transposable element detection and discovery for phylogenetic inference

Fritjof Lammers [1 2,*], Susanne Gallus [1], Axel Janke [1 2], Maria A Nilsson [1]

[1]Senckenberg Biodiversity and Climate Research Center, [2]Institute for Ecology, Evolution & Diversity, Goethe University Frankfurt, Frankfurt am Main, Germany

Abstract: Transposable Element (TE) insertions are a major cause of structural variation (SV) in mammalian genomes. Analyses of rare genomic changes are advantageous for phylogenetic inference and help to understand relationships among organisms. Furthermore, mutational effects of novel TE insertions can contribute to adaptation and speciation.
The diversity of TEs, and their continuous and irreversible integrations make them ideal phylogenetic characters that are almost free of homoplasy, however their detection is often laborious. Whole-genome sequencing allows to reconstruct phylogenies from TE insertion analyses at the genomic scale, but there are theoretical and practical challenges.
To enhance TE-based phylogenetics, we developed the "TeddyPi" pipeline (TE detection and discovery for phylogenetic Inference) that utilizes TE insertion calls from whole genome data of closely related species. By integrating different TE and SV callers and applying a strict filtering approach, TeddyPi can reliably and accurately detect informative TE insertions. TeddyPi generates high quality datasets of TE insertions for phylogenetic inference and other downstream genomic analyses.
In a pilot study, we extracted 85,534 TE insertions among six bear species with high confidence. Their analyses resolved the evolutionary history of bears for the first time solely with TE insertion data. In addition, a large amount of phylogenetic incongruence caused by incomplete lineage sorting and introgressive hybridization was detected by phylogenetic network analyses.
For any group of closely related species, TeddyPi opens up new possibilities for biologists to study phylogenies, evolutionary processes as well as rates and patterns of (retro-)transposition and structural variation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG5

Population genomics of Anolis carolinensis transposable elements: insertion polymorphisms are abundant but rarely approach fixation

Robert Ruggiero [1,*], Stephane Boissinot [1]

[1]Biology, NYU Abu Dhabi, Abu Dhabi, United Arab Emirates

Abstract: The importance of transposable element polymorphisms as a source of genetic variation in wild populations may be widely underappreciated. We are interested in understanding transposable element (TE) evolution and how TEs contribute to genome dynamics, so we conducted whole genome sequencing and analysis of 28 *Anolis carolinensis* genomes from five previously described populations. The *Anolis* lizard has a well-characterized genome and is known to harbor a diversity of active or recently active transposable elements, including several families of SINEs, LINEs, LTRs, and DNA transposons. Split-read based approaches were used to identify likely sites of polymorphic insertions for multiple subfamilies of each of these groups. We characterized the frequency of replicative success and found evidence for distinct rates of success across groups and subgroups. Additionally, we compared the frequencies of TE insertion polymorphisms to single nucleotide polymorphisms and confirmed prior expectations that demography has a significant effect on the persistence of TEs. We conclude that in these *Anolis* populations there is an abundance of low frequency TE insertion polymorphisms, representing a large and previously undescribed pool of transient structural variation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG3

Chromosomal inversions govern adaptation to altitude in honeybees

Andreas Wallberg [1], Caspar Schöning [2], Martin Hasselmann [3], Matthew Webster [1,*]

[1]Dept. of Medical Biochemistry & Microbiology, Uppsala University, Uppsala, Sweden, [2]Institute for Bee Research, Hohen Neuendorf, [3]Institute of Animal Science, University of Hohenheim, Hohenheim, Germany

Abstract: Populations of the honeybee *Apis mellifera* that inhabit the mountain forests of East Africa differ in behavior and morphology from those inhabiting the surrounding lowland savannahs, which likely reflects adaptation to these habitats. We performed whole genome sequencing on 39 samples of highland and lowland bees from two pairs of populations to determine their evolutionary affinities and identify the genetic basis of these putative adaptations. We find that in general, levels of genetic differentiation between highland and lowland populations are very low, consistent with them being a single panmictic population. However, we identify two loci on chromosomes 7 and 9, each several hundred kilobases in length, which exhibit near fixation for different haplotypes between highland and lowland populations. The highland haplotypes at these loci are extremely rare in samples from the rest of the world. Patterns of segregation suggest that recombination between haplotypes at each locus is suppressed, indicating that they comprise independent structural variants. The haplotype on chromosome 7 harbors nearly all octopamine receptor genes. These have a role in learning and foraging behavior in honeybees and are strong candidates for adaptation to highland habitats. Molecular analysis of a putative breakpoint indicates that it may disrupt the coding sequence of one of these genes. Divergence between the highland and lowland haplotypes at both loci is extremely high suggesting that they are ancient balanced polymorphisms that greatly predate divergence between the extant honeybee subspecies.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG1

Evolutionary genetics of structural variation in maize

Jeffrey Ross-Ibarra [1,*]

[1]Department of Plant Sciences, Center for Population Biology, and Genome Center, University of California, Davis, United States

Abstract: Flowering plant genomes vary in size more than 2000-fold, and more than half weigh in at 2Gb or larger. Yet much of what we know of the genomics of plant adaptation is from model systems with small, compact genomes that may in fact represent outliers among Angiosperms. Here I present our recent efforts to investigate the evolutionary genomics of structural variation in maize, an ancient polyploid with a 2.5Gb genome consisting primarily of transposable elements. Genome size varies across multiple altitudinal clines in both cultivated maize and natural populations of teosinte. Population genetic analysis suggests this variation has been molded by adaptive evolution, and greenhouse experiments and analysis of published data suggest a mechanistic link between genome size and flowering time. In addition to differences in gross genome size, both maize and teosinte exhibit widespread variation for a number of structural polymorphisms including copy number variation, inversion polymorphisms, and transposable elements. I discuss work documenting this variation – from copy number variation in a single hillside population of teosinte to range-wide patterns of inversion frequencies across cultivated maize – and highlight its potential role in shaping maize evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-173

Exceptional Gene Duplication Dynamics in the Cyanobacterium Acaryochloris: A Role for RecA Paralogs in Adaptation and Genome Instability

Scott Miller [1,*], Amy Gallagher [1], Emiko Sano [1]

[1]University of Montana, Missoula, United States

Abstract: Chlorophyll (Chl) *d*-producing cyanobacteria (*Acaryochloris* spp.) are a recently discovered and widely-distributed group of photosynthetic bacteria that uniquely possess this far-red light absorbing, structural relative of Chl *a* as the major pigment in photosynthesis. *Acaryochloris* genomes are both large (~8 Mbp) and contain an unusually high number of recent gene duplicates for bacteria. The origin of most duplicates appears to involve homologous recombination between different genetic elements, and the difference in gene duplication rate between *Acaryochloris* genomes is positively associated with copy number of the recombinase A (*recA*) gene. *Acaryochloris* genomes contain up to seven paralogous copies of *recA*, which is highly unusual, as this is a single copy gene in nearly all bacterial genomes. We show that these RecA paralogs are extremely diverse, differentially expressed and appear to have diverged in recombinase activity. The *Acaryochloris* system therefore presents a unique opportunity to gain novel insights on the consequences of RecA-mediated processes for genome evolution both within and between populations. We further show that duplicates that are retained are subject to strong purifying selection and that idiosyncratic duplicates in these genomes may contribute to adaptation to novel environments through positive dosage effects.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-222

Population genomics of transposable elements in a population of outcrossing Capsella grandiflora

Jasmina Uzunović [1,*], Stephen Wright [1], Emily Josephs [2], John Stinchcombe [1]

[1]Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada, [2]University of California, Davis, Davis, United States

Abstract: Transposable elements (TEs) make up a significant portion of eukaryotic genomes and thus, are a significant driving force of genome evolution. Ectopic recombination between TEs can cause genomic structural rearrangements and their insertion near genes can affect expression, regulation, and function. However, most of the present knowledge of TE dynamics comes from studying their distribution in reference genomes, making it difficult to untangle the most important evolutionary forces mediating TE copy number and genomic distribution. We characterised reference and non-reference TE insertions in a single population of 130 individuals of the obligate outcrosser, *Capsella grandiflora*. Most insertions are present in the population at a very low frequency (in only 1 individual), which is in line with the expected deleterious effects of TEs. The frequency of insertions was negatively correlated with distance to gene, as well as density of non-coding conserved elements. Furthermore, TEs which inserted near more highly expressed genes were at lower population frequency. In contrast, recombination rates were not a significant predictor of population frequencies. Taken together, these results suggest that the negative effects of TEs on gene regulation are more important in mediating TE loss as opposed to ectopic recombination.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG7

Evolution of satellite DNA loci in Drosophila melanogaster and the simulans clade

Emerson Khost [1,*], Danna Eickbush, Amanda Larracuente

[1]Biology, University of Rochester, Rochester, United States

Abstract: Many eukaryotic genomes are comprised of non-coding, repetitive DNA of unknown function. Satellite DNAs (satDNAs) are one such class of repetitive element, and are typically found in large tandem arrays in areas of reduced recombination (e.g. centromeres and telomeres). These arrays evolve rapidly: their genomic position can be vastly different even between closely related species, and within species array sizes can be polymorphic. However, their highly repetitive nature makes satDNA loci difficult to assemble, resulting in incomplete or uncertain assemblies in these regions. Here we overcome this limitation using single molecule real time sequencing from Pacific Biosciences (PacBio), which enables us to improve assembly of satDNA loci and examine their structure in fine detail. We use PacBio data from *Drosophila melanogaster* and from the three closely-related species of the *simulans* clade (*D. simulans, D. mauritiana,* and *D. sechellia*) to compare the distribution and composition of several families of complex satDNA. We estimate within-species variation in array organization and abundance using Illumina sequence data and quantify the influence of recombination, drift and selection on complex satellite arrays in the *simulans* clade.

Expanded summary*: One long-standing observation in biology is that the majority of the average eukaryotic genome is comprised primarily not of coding or regulatory sequence, but rather non-coding, highly repetitive DNA. One member of this class of DNA is satellite DNA (satDNA), which are simple or complex motifs repeated in long tandem arrays that can account for large fractions of the genome. SatDNA tend to accumulate in regions of low recombination, such as the heterochromatin of the centromere and telomere. They evolve extremely rapidly both within and between species: arrays expand, contract, or are lost entirely within very short evolutionary time scales. For example, the R*esponder (Rsp)* satellite localizes to chromosome 2R in *Drosophila melanogaster*, while in the closely related *simulans* clade *Rsp* has moved to the X chromosome in *D. simulans*, the 2nd and 3rd chromosomes in *D. sechellia*, and is absent at the cytological level in *D. mauritiana*. This represents a rapid structural rearrangement, as there only ~200,000 years separating the three *simulans* clade species. While there is theoretical work on the evolution of satDNA, technical constraints have ensured that until recently direct examination of these loci has been challenging. Due to the difficulty in studying highly repetitive loci, models of satDNA rely on assumptions regarding certain fundamental evolutionary parameters such as cross-over and mutation rates. In addition, most models treat satDNA as purely neutral or weakly deleterious. Their presence within a genome is usually attributed to purely selfish processes, with selection acting only to remove genomes with unacceptably large arrays. However, satDNA have important structural functions within the cell. For instance, satDNA loci affect chromosome pairing during achiasmatic meiosis and interact with proteins that maintain centromere identity. Furthermore, some satDNA are transcribed into short RNAs that associate with Piwi proteins and undergo ping-pong amplification, though what function these RNAs perform is currently uncertain. In addition, satDNA is hypothesized to play a role in preferential segregation of chromosomes during female meiosis, i.e. centromere drive. This would create opportunity for selection and genetic conflict, as centromere drive should be neutral in females but deleterious in males. Thus negative selection might not be the only form of selection acting on satDNA. Recent advances in sequencing technology now enable us to examine satellite DNA with high resolution and address some of these questions about their basic biology. We focus on two prevalent families of complex satellites in *D. melanogaster: Rsp* and members of the *1.688* family. Using single molecule real time sequencing from Pacific Biosciences (PacBio), we are able to assemble and detail the structure of highly repetitive satDNA regions that were previously inaccessible. We use these assemblies to compare how structure and sequence of satDNA loci change across *D. melanogaster* and the *simulans* clade. Within species, we have begun using population level Illumina sequencing data to determine satDNA sequence diversity and copy number variance. We can then combine this with estimates from mutation accumulation lines to model population genetic parameters such as the rate of exchange and ask whether these sequences are evolving non-neutrally.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-171

Passiflora plastomes are highly rearranged with multiple gene/intron losses, inversions and inverted repeat expansion/contraction

Bikash Shrestha [1,*], Samar Rabah [2], Nahid Hajrah [2], Ranai Makki [2], Hesham Alharby [2], Alawiah Alhebshi [2], Jamal Sabir [2], Lawrence Gilbert [1], Tracey Ruhlman [1], Robert Jansen [1 2]

[1]Department of Integrative Biology, University of Texas at Austin, Austin, United States, [2]Department of Biological Science, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract: Angiosperms lineages with highly rearranged plastid genomes (plastomes) often exhibit a syndrome of features, including biparental inheritance, plastome-genome incompatibility (PGI) and higher rates of nucleotide substitution. Evidence for biparental inheritance and PGI along with series of plastid gene and intron losses has been reported in few *Passiflora* species studied so far. However, knowledge of *Passiflora* plastome structure is limited because no genome sequence has been published. We present plastome sequences for 14 species from three subgenera: *Passiflora* (11), *Decaloba* (2), and *Astrophea* (1). Each subgenus has a distinct plastome organization. *Decaloba* plastomes are highly rearranged with substantial inverted repeat (IR) expansion. *Passiflora biflora* has the smallest plastome size at 139,263 kb with an IR expansion of ~7 kb IR, and *P. auriculata* with plastome size of 161,101 kb has a ~27 kb IR expansion. In contrast, *P. pittieri* (subg. *Astrophea*) has the largest plastome at 161,494 kb with both IR expansion (~5.4 kb) and contraction (~2.2 kb). Comparison of gene/intron losses, and inversions across a 64 plastid gene phylogeny identified shared and unique events associated with each subgenus, such as, loss of *rpl22* and *atpF* intron for the entire genus, loss of *rpl20, ycf1, ycf2 and clpP* introns in subgenera *Passiflora* and *Decaloba* and loss of *rps7* and *rpoC1* intron in *Decaloba.* Gene content, order, structure and overall plastome size of all the species from subg. *Passiflora* are similar. *Passiflora* plastome data is being utilized to investigate causes and consequences of genomic rearrangements and phylogenetic distribution of plastid inheritance.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-224

Extensive and variable repeat-mediated mitochondrial genome rearrangement in a genus of plants

Logan Cole [1,*], Jeffrey Mower [2], Jeffrey Palmer [1]

[1]Department of Biology, Indiana University, Bloomington, [2]Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, United States

Abstract: Angiosperm mitochondrial genomes evolve rapidly in sequence arrangement. This frequent rearrangement is generally attributed to intragenomic recombination mediated by repetitive sequences, however this not been investigated at a genome-scale nor has the question of whether rates of rearrangement vary significantly over time in angiosperms. Furthermore, it has proposed that these rearrangements may instead be the result of large-scale tandem duplication events followed by losses of gene duplicates. To investigate these questions, we reconstructed the rearrangement history of seven mitochondrial genomes in *Monsonia* (Geraniaceae). We show that rearrangements occur at generally very high rates in these genomes and also document significant variation in mitochondrial rearrangement rates. At the extreme, the hyperactive mitochondrial genome of *M. ciliata* has accumulated ~38 rearrangements during the ~1 million years since its common ancestry with *M. herrei*, whose rearrangement rate is much lower. We find that sites of mitochondrial DNA rearrangement are highly preferentially located in very close proximity to repeated sequences in *Monsonia*. This provides strong support for the hypothesis that rearrangements in angiosperm mitochondrial genomes occur largely through repeat-mediated recombination. Because there is little variation in the amount of repeat sequence across *Monsonia* species, we infer that the elevated rate of rearrangement in *M. herrei* probably reflects an elevated rate of mitochondrial recombination *per se*.

Expanded summary*: Extensive and variable repeat-mediated mitochondrial genome rearrangement in a genus of plants

Logan W. Cole, Jeffrey P. Mower, and Jeffrey D. Palmer

Though the extensive rearrangement of angiosperm mitochondrial genomes has been known about for over thirty years (Sederoff et al. 1981; Palmer and Herbon 1988), the evolutionary history of such rearrangements has remained largely unexplored. While it is known that the rate of mitochondrial genome rearrangement is much higher in plants than in animals, the rates in which these rearrangements occur, and how much these rates vary among different lineages, is poorly characterized. Because the mitochondrial genomes of plants are repeat-rich, the mechanism of mitochondrial genome rearrangement in plants has generally been assumed *prima facie* as repeat-mediated intragenomic recombination (Palmer and Herbon 1988), though little unambiguous evidence has been provided for this mechanism. Others have assumed plant mitochondrial genomes rearrange through mechanisms similar to those known from the mitochondrial genomes of animals (Darracq, Varre, and Touzet 2010), which have been shown to occur through large-scale tandem duplications followed by differential losses of the resulting gene copies (Boore 2000; Eberhard and Wright 2016).

In order to infer a mitochondrial genome rearrangement history, I reconstructed the ancestral arrangement at each internal node for the phylogeny of seven species in the genus *Monsonia*. Subsequently, I calculated rearrangement distances between all adjacent nodes, each of which corresponds to a branch of the phylogeny. With these distances, I calculated an estimate of the absolute rate of genome rearrangement by dividing the number of rearrangement events on each branch by its length in years. The resulting history of genome rearrangement revealed a rate of ~8.3 rearrangements per million years with at least ~10 fold variation in rates within the *Monsonia* lineage, which is only ~9 million years old. This extensive variation in rearrangement rates is contrasted with low variation in synonymous substitution rates (~1.4 fold) across the tree, with no apparent relationship between these rates.

I also used the distribution of repeat elements and rearrangement breakpoints to investigate the potential role of repeat-mediated intragenomic recombination in the rearrangement of these genomes. It was found that sites of mitochondrial DNA rearrangement are highly preferentially located in very close proximity to repeated sequences in *Monsonia*. Furthermore, in one species that belongs to a lineage that appears to have undergone extensive and recent recombination, *M. ciliata*, we find an especially strong association. This, together with the observation that tandem duplications are virtually absent in the *Monsonia* genomes, provides unambiguous support for the role of repeat-mediated recombination in the rearrangement of plant mitochondrial genomes over the other hypothesis

involving tandem duplication with differential gene loss, which makes no specific predictions with respect to the distribution of repeat sequences and rearrangement breakpoints.

My analyses provide substantial insight into to both the pattern and mechanism of rearrangement in plant mitochondrial genomes. This serves to further differentiate the assumptions made about their evolution from that of animal mitochondrial genomes. When compared to slowly evolving animal genomes, such as those in arthropods (Xu et al. 2006), the *Monsonia* mitochondrial genomes suggest that eukaryotic mitochondrial genomes vary in their rate of rearrangement by a factor of >20,000 fold. The distribution of repeats and rearrangement breakpoints also provides support for a starkly different molecular mechanism from that which underpins the rearrangement of animal mitochondrial genomes. This reinforces the understanding that the assumptions that we make and the models that we propose based on our understanding of animal mitochondrial genome evolution are insufficient for explaining the incredible variation that we observe in plant mitochondrial genomes.

Mitochondrial genome rearrangement is also important in the evolution of mating systems in plants, as it has been implicated as a primary molecular mechanism of shifts from monoecy to gynodioecy (Delph, Touzet, and Bailey 2007). It is thought that genome rearrangement results in the formation of chimeric genes that are responsible for cytoplasmic male sterility (Hanson 1991; Mower et al. 2012) which leads to the coexistence of female and hermaphroditic plants within populations. The apparent discordance between synonymous substitution rates and rates of genome rearrangement in *Monsonia* genomes suggests that differential selection on mating system could be occurring across the lineage, resulting in changes in the intrinsic rate of rearrangement. Conversely, differential rates of repeat-mediated recombination could alter the rate at which these chimeric genes are formed, impacting the evolution of mating systems.

This work has significant implications for molecular- and organism-level processes at population- and phylogenetic-scales. It reveals the vast differences in pattern and mechanism between plant and animal mitochondrial genome rearrangement that are the result of over a billion years of evolution. Furthermore, I observe surprisingly great differences in rearrangement rates among closely related species, which may explain variation in mating systems frequencies within those species.

REFERENCES

Boore, J.L. 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostrome animals In D. Sankoff and J. Nadeau, editors, *Comparative Genomics*. Kluwer. **Darracq, A., Varré J.S., and Touzet, P. 2010.** A scenario of mitochondrial genome evolution in maize based on rearrangement events. *BMC Genomics* 11:233. **Delph, L. F., Touzet, P., and Bailey, M.F. 2007.** Merging theory and mechanism in studies of gynodioecy. *Trends in Ecology and Evolution* 22(1): 17-24. **Eberhard, J.R. and Wright T.F.** 2016. Rearrangement and evolution of mitochondrial genomes in parrots. *Molecular Phylogenetics and Evolution* 94: 34-46. **Hanson, M.R.** 1991. Plant mitochondrial mutations and male sterility. *Annual Review of Genetics* 25: 461-486. **Mower, J.P., Case, A.L., Floro, E.R., and Willis, J.H.** 2012. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. *Genome Biology and Evolution* 4: 670-686. **Palmer, J.D. and Herbon, L.A.** 1988. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution* 28: 87-97. **Sederoff, R.R., Levings III, C.S., Timothy, D.H., and Hu, W.W.L.** 1981. Evolution of DNA sequence organization in mitochondrial genomes of *Zea*. *Proceedings of the National Academy of Sciences USA* 78(10): 5953-5957. **Xu, W., Jameson, D., Tang, B., and Higgs, P.G.** 2006. The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes. *Journal of Molecular Evolution* 63: 375-392.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-204

Ongoing loss of duplicates created by whole-genome duplication in natural populations of Paramecium species

Parul Johri [1,*], Michael Lynch [1]

[1]Biology, Indiana University, Bloomington, United States

Abstract: Whole genome-duplications (WGDs) have occurred in the ancestors of many eukaryotes. Several studies have suggested that duplicates created by WGD might be retained initially as a result of dosage compensation, i.e. selection acts to maintain the sum of expression of the two copies, in order to maintain the stoichiometry of the protein product relative to other interacting proteins. We evaluate this hypothesis in a group of five species of *Paramecium*- *P. tetraurelia, P. biaurelia* and *P. sexaurelia* that maintain ~50% of the duplicates created by the most recent WGD, and two outgroup species- *P. caudatum* and *P. multimicronucleatum*, that diverged before the WGD. We identified copy number variants to detect duplicate presence-absence polymorphisms among 10-13 individuals within each species and characterized the patterns of polymorphism and divergence genome-wide. We find that genes that tolerate more non-functional mutations in populations of outgroup species, and are thus more dosage sensitive, are more likely to be lost post-WGD. Genes under stronger purifying selection in the outgroups are more likely to be retained post-WGD, consistent with the idea that genes with more interacting partners are less likely to be lost. In addition, genes still maintained as duplicates, have significantly lower levels of diversity in their upstream intergenic region, in contrast to genes present in single-copy. And non-synonymous to synonymous diversity is correlated between paralogs within species, suggesting similar selective constraints acting on paralogs. Together these results emphasize the conservation of expression regulation and function of maintained duplicates, as expected under the dosage compensation model.

Expanded summary*: Gene duplication is an important source of evolutionary novelty, resulting in new gene functions, expansion of gene families and contributing to the emergence of new species. New gene copies can arise either via segmental duplications, in which a chromosomal segment containing one or more genes is duplicated, or via whole genome duplications (WGDs). Segmental duplications have been found to occur at high rates in most eukaryotic genomes, while WGDs have occurred in the ancestors of many eukaryotes: e.g. *Saccharomyces cerevisiae*, *Xenopus laevis*, teleost fish and *Arabidopsis thaliana* and nearly all land plant genomes appear to have experienced at least one WGD. Although most new gene duplicates are rapidly lost from a population, a number of duplicates are preserved either exactly or are retained with a change in function. However, it is not well understood why certain genes are more prone to retention than others and what evolutionary forces are responsible for long-term retention of duplicates. Given that a number of polyploid species retain 25-75% of duplicates generated by the most recent WGD, a significantly higher proportion than those generated by segmental duplications, it has been proposed that maintaining stoichiometry of protein products allows for a longer retention of duplicates immediately post-WGD (dosage compensation model).

To elucidate the evolutionary forces influencing the fate of gene duplicates, we need to understand the evolutionary processes operating in the early post-duplication stages in populations. Studies of copy number variants in humans and *Drosophila* have shed light on evolutionary forces responsible for their preservation and elimination. Similar studies are required to understand the population dynamics of gene duplicates following WGD and the early factors that impact their long-term retention. We propose to do this by comparative population genomics of a complex of three *Paramecium aurelia* species (with WGDs), and two outgroup species (without WGDs).

We identified copy number variants (CNVs) to detect duplicate presence-absence polymorphisms and conducted analyses of polymorphism and divergence across all genes for all five species. We find that genes that tolerate more non-functional mutations in populations of outgroup species, and are thus more dosage sensitive, are more likely to be lost post-WGD. Using neutrality index of genes, we show that genes under stronger purifying selection are more likely to be retained post-WGD and continue to be under stronger purifying selection, suggesting that dosage compensation plays a major role in retention of duplicates created by WGD. Moreover, if most duplicates have been preserved by dosage compensation, they are expected to have the same ancestral function. Interestingly, retained duplicate copies have significantly lower levels of diversity in their upstream intergenic region, in contrast to the single-copy genes. A strong correlation of non-synonymous to synonymous diversity between paralogs within species suggests similar selective constraints on paralogs, and thus emphasizing the conservation of regulation and function of maintained duplicates, as expected under the dosage compensation model.

Currently we are characterizing potential candidates undergoing non-functionalization in populations, in order to better understand the mechanism and rate of duplicate loss. This would be the first study to use population-genomic data to understand the mechanism of duplicate-gene evolution across multiple closely-related taxa, especially duplicates that resulted from whole-genome duplications.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-217

Disulfide bonds enable accelerated protein evolution

Felix Feyertag [1,*], David Alvarez-Ponce [1]

[1]Biology, University of Nevada at Reno, Reno, United States

Abstract: The different proteins of any proteome evolve at enormously different rates. What factors contribute to this variability, and to what extent, is still a largely open question. We hypothesized that disulfide bonds, by increasing protein stability, should make proteins' structures relatively independent of their amino acid sequences, thus acting as buffers of deleterious mutations and enabling accelerated sequence evolution. In agreement with this hypothesis, we observed that membrane proteins with disulfide bonds evolved 88% faster than those without disulfide bonds, and that extracellular proteins with disulfide bonds evolved 49% faster than those without disulfide bonds. In addition, genes encoding proteins with disulfide bonds exhibit an increased likelihood of showing signatures of positive selection. Multivariate analyses indicate that the trend is independent of a number of potentially confounding factors. The effect, however, is not observed among the longest proteins, which can become stabilized by mechanisms other than disulfide bonds.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

OT-EG14

Detecting and understanding the action of natural selection within segmental duplications in primates

Josephine Daub [1,*], Arcadi Navarro [1]

[1]Institute of Evolutionary Biology , Universitat Pompeu Fabra - CSIC, Barcelona, Spain

Abstract: Many of the differences in traits that we observe between species are likely the result of adaptations by means of positive selection. However, when comparing humans and their close relative the chimpanzee, only a few examples of such adaptations are detected at the genome level. It is not unlikely that such signals of selection could be found in often overlooked regions of the genome, namely segmental duplications and non-coding functional genomic elements. Both fields pose technical challenges, and it is for that reason that the number of selection scans that target these genomic areas has been low up until recently. With the rise of new sequencing technologies, such as long-read sequencing, and the advance in knowledge of non-coding elements, some of the standing questions can now be answered. I will present a robust approach to detect positive selection in coding and non-coding regions on a genome wide scale, which is specifically suited to detect selection in segmental duplications. In short, this method is based on the creation of a so-called consensus sequence after alignment of reads of multiple species to genes and their paralogs in a reference genome. By comparing the rate of variation in functional and neighboring regions, positive selection can be inferred. Applying this method on Pacbio datasets of human, chimpanzee, gorilla and macaque, we performed selection scans to gain more insight in the role of natural selection within segmental duplications in primates.

Expanded summary*: Many of the differences in traits that we observe between species are likely the result of adaptations by means of positive selection. However, when comparing humans and their close relative the chimpanzee, only a few examples of such adaptations are detected at the genome level. It is not unlikely that such signals of selection could be found in often overlooked regions of the genome, namely segmental duplications and non-coding functional genomic elements. Both fields pose technical challenges, and it is for that reason that the number of selection scans that target these genomic areas has been low up until recently. With the rise of new sequencing technologies, such as long-read sequencing, and the advance in knowledge about non-coding elements, some of the standing questions can now be answered.

I will present a robust approach to detect positive selection in coding and non-coding regions on a genome wide scale, which is specifically suited to detect selection in segmental duplications. It is an extension of an earlier method developed in our lab which has been used to successfully detect several gene families that were possibly affected by positive selection[1]. In short, this method is based on the creation of a so-called consensus sequence after alignment of reads of multiple species to genes and their paralogs in a reference genome. By comparing the rate of variation in functional and neighboring regions, positive selection can be inferred.

Using an extended version of the method, which can target non-coding genomic elements and allows for the use of long-read sequences, we performed genome wide selection scans using human, chimpanzee, gorilla and macaque Pacbio data sets. I will present the results of these scans and compare them with earlier findings produced by a previous version of the method.

The extensive study of positive selection in segmental duplications of primates will improve our understanding of the relations between selection, molecular diversity, structure and function in the genomes of humans, primates and other species, and it will thus help to advance evolutionary biology and, potentially, medicine.

Keywords: Positive selection, segmental duplications, non-coding elements, primates

[1] Lorente-Galdos, B. et al. Accelerated exon evolution within primate segmental duplications. Genome Biology 14, R9 (2013).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-207

Copy number variation and a role in vision for CRAL-TRIO domain genes in Heliconius melpomene

Aide Macias Muñoz [1,*], Kyle McCulloch [1], Adriana Briscoe [1]

[1]Ecology and Evolutionary Biology, University of California, Irvine, Irvine, United States

Abstract: The CRAL-TRIO domain is a structural region common to proteins that transport hydrophobic tocopherols. In vertebrates (CRALBP) and *Drosophila* (PINTA) CRAL-TRIO domain containing proteins transport vitamin A-derived chromophores that are necessary for vision. Members of the CRAL-TRIO domain gene family have undergone lineage-specific duplications in insects, and an expansion in Lepidoptera. Lepidopterans have twice as many CRAL-TRIO domain genes compared to other insects. However, there is no lepidopteran ortholog of *pinta*. We aimed to 1) characterize the molecular evolution of the CRAL-TRIO domain gene family and to 2) identify a candidate gene for chromophore transport in butterflies. By searching a *de novo* transcriptome and reference genome, we found 43 CRAL-TRIO domain genes in a butterfly species, *Heliconius melpomene*. A phylogeny revealed two duplication events and an expansion of these genes in *H. melpomene*. 36 of the CRAL-TRIO domain genes were located in tandem on 3 chromosomes. We used 18 resequenced genomes from 4 subspecies to detect copy number variation of 32 CRAL-TRIO domain genes. We also performed differential expression analysis using RNA-Seq from the heads, antennae, legs and mouthparts of *H. melpomene* to identify a candidate CRAL-TRIO domain gene, *Hme CTD31,* upregulated in heads. RT-PCR confirmed that *Hme CTD31* is expressed in the retina rather than the brain. Furthermore, immunohistochemistry showed that the Hme CTD31 protein is found in primary and secondary pigment cells. The CRAL-TRIO domain gene family is likely evolving by tandem duplications and a member of this family potentially functions in butterfly visual pigment transport.

Expanded summary*: Gene duplication events play a role in evolution by giving rise to genetic material that can acquire new functions. Many gene family expansions are the result of gene duplications. Lineage-specific gene family expansions are hypothesized to be a mechanism by which eukaryotic species can adapt and diversify. Members of the CRAL-TRIO domain-containing gene family have undergone lineage-specific duplications in insects. The CRAL-TRIO domain is a N-terminal structural region common to several proteins that bind and transport tocopherols. Vertebrate (CRALBP) and *Drosophila* (PINTA) proteins with a CRAL-TRIO domain are known to transport vitamin A-derived chromophores that bind to opsin proteins and are necessary for phototransduction. In a lepidopteran-specific expansion, CRAL-TRIO duplication events have resulted in twice as many CRAL-TRIO domain genes compared to other insects. However, there is no lepidopteran ortholog of *pinta*. This project aims to 1) characterize the molecular evolution of the CRAL-TRIO domain gene family and to 2) identify a candidate gene for chromophore transport in butterflies.

I used RNA-Seq libraries from the heads, antennae, legs and mouthparts of the butterfly *Heliconius melpomene* to build a *de novo* transcriptome. I searched the *de novo* transcriptome and a reference genome to identify a total of 43 CRAL-TRIO domain genes. I discovered two duplication events and an expansion of CRAL-TRIO domain genes in *H. melpomene*. Most of these genes were located in tandem. As this suggested evolution by duplication events, I used 18 resequenced genomes from 4 subspecies to search for copy number variation (CNV) between individuals. I found potential CNV in 32 of the 43 CRAL-TRIO domain genes.

In order to identify a candidate gene involved in chromophore transport, I performed differential expression analysis to find genes upregulated in head tissue relative to antennae, legs, and mouth parts. Head upregulated genes were annotated with *Drosophila* gene ontology terms and enriched for function. The top three annotation clusters had a function in vision and the genes grouped in these clusters encode proteins involved in phototransduction. I found one CRAL-TRIO domain gene, *Hme CTD31,* upregulated in heads. RT-PCR confirmed that *Hme CTD31* is expressed in the retina rather than the brain. Furthermore, immunohistochemistry shows that the protein Hme CTD31 is expressed in primary and secondary pigment cells and also in trachea cells making Hme CTD31 a candidate chromophore or filtering pigment transporter.

Significance: This research project makes use of large-scale data sets to increase our knowledge of the molecular evolution, expression patterns, and CNV of CRAL-TRIO domain genes in butterflies. My findings support a role of tandem duplications and CNV in gene family expansions and evolution. The adaptive significance of CNV is still under investigation. When it comes to *Heliconius* butterflies, it is posited that CNV may be contributing to the speciation of this genus that has undergone a radiation in Central and South America.

Furthermore, the functions of CRAL-TRIO domain proteins in organisms aside from humans and *Drosophila* remain largely unexplored. In vertebrates and *Drosophila*, a member of this gene family transports the chromophore molecule. The CRAL-TRIO domain gene family has undergone an expansion in Lepidoptera where not much is known about the genes involved in chromophore or filtering pigment transport. I identify a CRAL-TRIO domain gene expressed in butterfly retina with putative function in butterfly visual systems providing an advance in the field. These results suggest that CRAL-TRIO domain genes have conserved function in vision across species. In addition, lepidopteran visual systems have evolved through duplications of opsin genes and CRAL-TRIO domain genes seem to have evolved through similar mechanisms. Exploring the role of these genes in vision may paint a better picture of how complex traits like vision evolve.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-216

Plastome sequencing of 10 non-model crop species uncovers a large insertion of mitochondrial DNA in cashew (Anacardium, Anacardiaceae)

Chaehee Lee [1,*], Samar Rabah [2], Nahid Hajrah [3], Ranai Makki [3], Hesham Alharby [3], Alawiah Alhebshi [3], Jamal Sabir [3], Meshaal Sabir [3], Robert Jansen [1 3], Tracey Ruhlman [1]

[1]Department of Integrative Biology, The University of Texas at Austin, Austin, United States, [2]Department of Biological Sciences, Faculty of Science, [3]Genomics and Biotechnology Section and Reserach Group, Department of Biological Sciences, Faculty of Science, King Abdulaziz University, jeddah, Saudi Arabia

Abstract: In plant evolution, intracellular gene transfer (IGT) is a prevalent and ongoing process. While nuclear and mitochondrial genomes are known to integrate foreign DNA via IGT, plastomes of land plants have resisted both horizontal gene transfer (HGT) and IGT. Only recently has foreign DNA incorporation into plastomes been uncovered, facilitated in part by the availability of complete mitogenome sequences. In this study, we completed plastomes for l0 non-model crop species. We describe structural variations including gene and intron content, inversions and both expansion and contraction of the inverted repeat (IR). In one lineage, the orchard tree cashew (*Anacardium occidentale*), we report the insertion of a ~6.7 kb fragment of mitochondrial DNA in the plastome IR. Blastn analyses returned high identity hits to plant mitogenomes including genic and intergenic sequences. Illumina read mapping to the *A. occidentale* assembly and PCR amplification of boundary sequences confirmed plastome insertion in the IR. Based on our PCR survey of eight *Anacardium* species, four shared the insertion suggesting that this event occurred less than 5 MYA within the genus. Our study extends the observation of mitochondrial to plastome IGT to include long-lived tree species. While previous studies have suggested possible mechanisms facilitating IGT to the plastome, more examples of this phenomenon will be required before a common, or variable, mechanism(s) can be elucidated.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-209

New Drosophila reference genomes unveil previously hidden genetic variation

Mahul Chakraborty [1,*], J. J. Emerson [2], Anthony D. Long [2], Stuart J. McDonald [3]

[1]Department of Ecology and Evolution, [2]Department of Ecology and Evolutionary Biology, University of California Irvine, Irvine, [3]Molecular Biosciences, The University of Kansas, Lawrence, United States

Abstract: Accurate characterization of genetic variation is essential for understanding phenotypic evolution. While high throughput short reads are well-suited to discovery of single nucleotide and small indel polymorphisms, large scale structural mutations (e.g. duplications, deletions, insertions, etc.) are often overlooked by such methods. Unfortunately, such mutations often play pivotal roles in genome evolution and the genetic basis of disease. Thus, our perception of structural genetic variation is severely limited by the current methods. To overcome these limitations, we resequenced the founder strains of the Drosophila Synthetic Population Resources (www.flyrils.org) using Pacific Biosciences long reads. To shed light on hidden structural variants (SVs), we constructed de novo assemblies for each of these strains. Notably, completeness and contiguity of the assemblies are comparable to or better than the current release of the reference strain, with the majority of the genome represented by contiguous sequences (contigs) measuring 20Mb or longer. Comparisons of these assemblies revealed ubiquitous duplicates, transposon insertions, and inversions, revealing the dynamic nature of genome structure. A large number of these SVs, nearly 50% of which were previously unknown, contribute to gene structure polymorphism, expression level variation, and phenotypic adaptations, several of which we describe in detail.

Expanded summary*: Although structural variants (sequence duplication, insertion-deletion, and transpositions) contribute to diseases and other phenotypic changes, current high throughput short read sequencing methods often fail to identify such mutations in a genome. Consequently, it has been predicted that many hidden SVs represent missing candidates in GWAS and QTL experiments that have failed to yield a causal mutation. While these hidden variants can be detected by aligning two genome assemblies in which all SVs have been assembled correctly, virtually all genomes used for genotyping purposes are assembled or genotyped using short reads and therefore they are missing the genomic regions enriched with SVs. My postdoctoral research aims at addressing these challenges by constructing extremely contiguous genome assemblies with PacBio long reads.

However, PacBio long reads being expensive, to reduce the burden of PacBio sequencing cost, I wrote a new metassembly algorithm that lowered the PacBio read coverage requirements for generating extremely contiguous assemblies. The metassembler, which is implemented in a freely available software called *quickmerge*, was used to assemble a new *D. melanogaster* reference genome (called A4) that is more contiguous (50% of the genome is contained within sequences of length 22.3Mb (N50) or longer) than the current version (release 6) of the *D. melanogaster* reference genome (N50 = 21.3Mb). Next, using a SV (CNVs, TE insertions-deletions, inversions) detection pipeline that I developed, called SVMU (Structural Variants from Mummer) I created a comprehensive SV map of the A4 genome. This comprehensive SV map revealed novel gene duplicates implicated in toxin resistance, cold adaptation, mating behavior, and olfaction. Furthermore, the SV map revealed a staggering amount of gene structure polymorphism caused by TE insertions. Surprisingly, nearly 50% of these mutations were invisible to high coverage illumina paired end reads. The fitness and phenotypic effects of a subset of these previously hidden gene duplicates are currently being characterized using CRISPR.

Motivated by the discovery of novel structural variants from A4-ISO1 genome alignment, I began a collaboration with Dr. Tony Long to resequence all 15 founders of the mapping panel known as Drosophila Synthetic Population Resources (www.flyrils.org) with PacBio long reads. The goal of this project was to generate a comprehensive SV map for each genome and then integrate the genotypes into the existing genotype database to re-examine the missing candidates in previously conducted QTL studies. So far I have finished assembling all genomes for this project, with each genome reaching the same level of completeness as A4 but with half the coverage, partly due to the improvements I made to the metassembly algorithm since the assembly of A4. Preliminary analysis of these genomes have uncovered novel candidate structural mutations for nicotine resistance, caffeine resistance, and chemotherapy drug resistance. These previously unknown SVs suggest that a significant number of candidate structural variants in *Drosophila* are invisible to the current illumina short reads based CNV or TE insertion genotyping methods. These newly discovered SVs also suggest that hidden SVs are often the missing candidates in QTL and GWAS studies. Based on these results from DSPR, I expect that an explosion of discovery of such novel candidate SVs in other organisms, including humans, will soon follow.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-205

Variable rates of satellite DNA gains across the Drosophila phylogeny

Kevin Wei [1 2,*], Sarah Sander [2], Daniel Barbash [2], Andrew Clark [2]

[1]Integrative Biology, Univerisyt of Califonia Berkeley, Berkeley, [2]Molecular Biology and Genetics, Cornell University, Ithaca, United States

Abstract: Tandemly-repeated DNA elements known as satellite DNA occupy significant portions of eukaryotic genomes, despite their propensity to cause genomic instability through non-allelic exchange. They are primarily found in the repressive heterochromatin near the centromeres and telomeres, and on the Y where they accumulate as the chromosome degenerates. Interestingly, the types and abundances of satellites often vary dramatically between closely related species, suggesting that they turn over rapidly. However, limited sampling has prevented detailed understanding of their evolutionary dynamics. Here, we comprehensively characterize simple satellites from whole-genome sequences generated from males and females of nine *Drosophila* species, spanning 40 million years of evolution. The libraries were prepared without PCR amplification which we show better captures satellite quantities. We identified few satellites shared across species, consistent with previous descriptions of their rapid evolution. Using a maximum parsimony framework, we surprisingly find that almost all interspecific differences are due to lineage-specific gains with losses being rare. Furthermore, rates of change are highly variable among lineages, with the melanogaster complex species having the highest rates of gain. Examining *D. melanogaster* and *D. virilis*, two of the most satellite-rich species, we find that younger satellites can be traced through single mutations to older satellites, indicating that novel satellites primarily emerge from pre-existing ones. These results reveal the complexities and variability of satellite DNA evolution and provide insights into the molecular and evolutionary mechanisms driving their change.

Expanded summary*: Most eukaryotic genomes contain large quantities of tandemly-repeating, non-coding sequences known as satellite DNA. These sequences, along with transposable elements, are often described as genomic parasites since they increase in copy number often at the expense of host fitness [1]. Due to their repetitive nature, satellite DNA have high capacity to induce homology-directed crossover events which on the one hand mediates expansions and contractions through unequal crossovers [2], but on the other creates chromosome rearrangements and genome instability [3]. Consequently, these elements are mostly sequestered in repressive chromatin known as heterochromatin typically found around centromeres and telomeres where recombination and transcription are minimal.

Despite their deleterious potential, satellite DNA can also adopt crucial cellular functions; for example, they recruit the centromeric histone H3 variant to form centromeres in many taxa. Furthering the interest, satellite DNA content between closely related species is often poorly conserved, leading to the notion that they have rapid turnover. Comparisons between *Drosophila melanogaster* and its sister species *D. simulans* revealed that not only do they differ in the total quantities (20% vs 5%, respectively), they also harbor distinct types [4,5]. This difference has even been implicated in hybrid incompatibility, indicating that the divergence of satellite DNA can play an important role in speciation [6]. Moreover, their centromeric and telomeric functions have also led to speculations that satellite DNA can harbor or become meiotic drivers and their rapid change is the result of recurrent replacement by emerging drivers [7,8].

Given the many fascinating facets of satellite DNA, the mechanisms driving their evolution remain unclear and poorly explored. With my co-advisors Daniel Barbash and Andrew Clark, my graduate research focuses on understanding how repetitive sequences evolve in Drosophila. I took a three pronged approach: characterize satellite DNA variation within and between populations of *D. melanogaster*, examine divergence across multiple Drosophila species, and determine whether their rapid turnover resulted from meiotic drive. To characterize satellite DNA, I circumvented the difficulties repetitive and low complexity sequences typically present to sequence analyses by developing k-Seek, a method that identifies *de novo* and quantifies simple-sequence repeats from short read whole genome sequences [9]. I applied k-Seek on 84 inbred *D. melanogaster* lines derived from populations collected worldwide to determine how satellite DNA evolve within species. In addition to all the known ones, many novel satellites were identified, two of which are, interestingly, population-specific. I further found that the abundance of many satellites is correlated across populations, suggesting concerted growth and/or retraction.

To understand satellite DNA evolution at a greater time scale, I used k-seek to characterize satellite DNA contents of nine *Drosophila* species, spanning over 40 million years of evolution. Consistent with the notion of rapid turnover, I found few satellites shared between species. Surprisingly, interspecific differences are predominantly driven by rapid gains of novel satellites in some lineages, including the *melanogaster* complex, while others like *D. pseudoobscura* and *D. persimilis* appears to have acquired few to no satellites over long evolutionary time. This suggests that rapid turnover is not a universal characteristic of satellite evolution and that some genomes may be sensitive to satellite abundance. I will present these findings at SMBE 2017 Austin.

To explore the possibility for satellite DNA as meiotic drivers, we focused on female meiosis, where genetic loci can exploit the asymmetric cell division during oogenesis, thus increasing their transmission rate into the oocyte and fixation rate in the population. While centromeres and telomeres are likely hotspots for such drivers to accumulate, few have ever been identified [10]. Part of the difficulty lies in detecting weak drivers even though they are expected to be more prevalent as they are slower to reach fixation [11]. I developed a method that sensitively detects minor distortions to Mendelian ratio from sequencing of large embryo pools. Using this method, we found no evidence that *Drosophila* telomeres which comprise large tandem arrays of the retrotransposon *HeT-A* affect segregation fidelity. Curiously, we identified a candidate locus linked to the centromere that distorts segregation frequency. The satellite composition at/near this centromere is now under exploration. Using this method we further tested whether the aforementioned population-specific satellites can cause meiotic drive. Interestingly, we found that they do show elevated segregation rates, however, additional assays to rule out confounding effects are required and underway.
The completion of my Ph.D. has heavily relied on a plethora of genetic, genomic, and evolutionary approaches. Resources provided by the SMBE including stellar publications in MBE and GBE, annual SMBE meetings, and feedback from members have been instrumental in my progress. As I intend on pursuing an academic career in evolutionary genomics and genetics, I expect that SMBE will continue to play a pivotal role in my research endeavors. Reciprocally, I expect that our methods and findings are of wide interest to the community. Hopefully with the support of SMBE, I will continue to contribute and become a productive member of this great community.

References:

1. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. Nature. 1980;284: 601–603.
2. Smith GP. Evolution of repeated DNA sequences by unequal crossover. Science. 1976;191: 528–535.
3. Peng JC, Karpen GH. Heterochromatic genome stability requires regulators of histone H3 K9 methylation. PLoS Genet. 2009;5: e1000435.
4. Lohe AR, Brutlag DL. Multiplicity of satellite DNA sequences in Drosophila melanogaster. Proc Natl Acad Sci U S A. 1986;83: 696–700.
5. Lohe AR, Brutlag DL. Identical satellite DNA sequences in sibling species of Drosophila. J Mol Biol. 1987;194: 161–170.

6. Ferree PM, Barbash DA. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila. PLoS Biol. 2009;7: e1000234.
7. Zwick ME, Salstrom JL, Langley CH. Genetic variation in rates of nondisjunction: association of two naturally occurring polymorphisms in the chromokinesin nod with increased rates of nondisjunction in Drosophila melanogaster. Genetics. 1999;152: 1605–1614.
8. Malik HS. The centromere-drive hypothesis: a simple basis for centromere complexity. Prog Mol Subcell Biol. 2009;48: 33–52.
9. Wei KH-C, Grenier JK, Barbash DA, Clark AG. Correlated variation and population differentiation in satellite DNA abundance among lines of Drosophila melanogaster. Proc Natl Acad Sci U S A. 2014;111: 18793–18798.
10. Fishman L, Willis JH. A novel meiotic drive locus almost completely distorts segregation in mimulus (monkeyflower) hybrids. Genetics. 2005;169: 347–353.
11. Thomson GJ, Feldman MW. Population genetics of modifiers of meiotic drive. III. Equilibrium analysis of a general model for the genetic control of segregation distortion. Theor Popul Biol. 1976;10: 10–25.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-225

Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure.

Tracey Ruhlman*, Jin Zhang, John C Blazier, Jamal Sabir, Robert Jansen

Abstract: In the plastid genomes (plastomes) of land plants, the variable orientation of the single copy regions relative to the inverted repeat (IR) was hypothesized to occur through intramolecular recombination in a circular, single unit-genome. In fact, inversion of the single copy regions more likely occurs through recombination-dependent replication (RDR) of linear plastome templates. If RDR can be primed through both intra- and intermolecular recombination, then this mechanism could not only create inversion isomers of so-called single copy regions, but also an array of alternative sequence arrangements. We used Illumina paired-end and PacBio single-molecule real-time (SMRT) sequences to characterize repeat structure in the plastome *Monsonia emarginata* (Geraniaceae). We used OrgConv and inspected nucleotide alignments to infer ancestral nucleotides and identify gene conversion among repeats and mapped long (>1 kb) SMRT reads against the unit-genome assembly to identify alternative sequence arrangements. Although *M. emarginata* lacks the canonical IR, we found that large repeats (>1 kilobase; kb) represent ~22% of the plastome nucleotide content. Among the largest repeats (>2 kb) we identified GC-biased gene conversion and mapping filtered, long SMRT reads to the *M. emarginata* unit-genome assembly revealed alternative, substoichiometric sequence arrangements. We offer a model based on RDR and gene conversion between long repeated sequences in the *M. emarginata* plastome, and provide support that both intra-and intermolecular recombination between large repeats, particularly in repeat-rich plastomes, varies unit-genome structure while homogenizing the nucleotide sequence of repeats.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-223

Rapid interspecific evolution of sex chromosome amplicons suggests candidates for meiotic drive and hybrid sterility within Felidae

Wesley Brashear [1,*], Terje Raudsepp [1], William Murphy [1]

[1]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, United States

Abstract: The ampliconic regions of sex chromosomes are among the most rapidly evolving in the mammalian genome. The highly repetitive nature of these regions makes them exceedingly difficult to sequence using modern short-read WGS approaches, and, as a result, has led to their absence from most current genome assemblies. This is surprising as these regions possess characteristics that potentially have major implications in meiotic drive systems, male fertility, and inter-specific hybridization. We have used an integrated approach combining chromosome-specific cDNA capture, RNA-Seq, FISH, BAC sequencing, and whole genome analyses to examine sex chromosome evolution within the family Felidae. We have identified several novel genes located on the X chromosome that show considerable copy number variation between species. Furthermore, several of these genes are located in close proximity to mapped hybrid sterility loci, suggesting plausible candidate genes for further interrogation. We have also sequenced a large representative portion of the domestic cat ampliconic MSY which has revealed a complex pattern of historic and ongoing duplication events and structural rearrangements. Comparative fluorescent in-situ hybridization and WGS data revealed a high degree of conservation of genic content, but variable copy number and arrangement, within the MSY across the cat family. These results reveal promising avenues for investigating the role of sex chromosomes in felid reproductive isolation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary genomics of structural variation

POA-219

Expressed Structurally-stable Inverted Duplicates in Mammalian Genomes as Functional Noncoding Elements

Zhen-Xia Chen*, Brian Oliver, Yong Zhang, Ge Gao, Manyuan Long

Abstract: Inverted duplicates are a type of repetitive DNA motifs consist of two copies of reverse complementary sequences separated by a spacer sequence. They can lead to genome instability and many may have no function, but some functional small RNAs are processed from hairpins transcribed from these elements. It is not clear whether the pervasive numbers of such elements in genomes, especially those of mammals, is the result of high generation rates of neutral or slightly deleterious duplication events or positive selection for functionality. To test the functionality of intergenic inverted duplicates without known functions, we used mirror duplicates, a type of repetitive DNA motifs with few reported functions and little potential to form hairpins when transcribed, as a nonfunctional control. We identified large numbers of inverted duplicates within intergenic regions of human and mouse genomes, as well as 19 other vertebrate genomes. Structure characterization of these inverted duplicates revealed higher proportion to form stable hairpins compared to converted mirror duplicates, suggesting that inverted duplicates may produce hairpin RNAs. Expression profiling across tissues demonstrated that 7.8% of human and 5.7% of mouse inverted duplicates were expressed even under strict criteria. We found that expressed inverted duplicates were more likely to be structurally stable than both unexpressed inverted duplicates and expressed converted mirror duplicates. By dating inverted duplicates in the vertebrate phylogenetic tree, we observed higher conservation of inverted duplicates than mirror duplicates. These observations support the notion that expressed inverted duplicates may be functional through forming hairpin RNAs.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES2

Active Interaction Mapping reveals the hierarchical organization of autophagy

Trey Ideker*

Abstract: We have developed a general progressive procedure, Active Interaction Mapping, to guide assembly of the hierarchy of functions encoding any biological system. Using this process, we assemble an ontology of functions comprising autophagy, a central recycling process implicated in numerous diseases. A first-generation model, built from existing gene networks in Saccharomyces, captures most known autophagy components in broad relation to vesicle transport, cell cycle and stress response. Systematic analysis identifies synthetic-lethal interactions as most informative for further experiments; consequently, we saturate the model with 156,364 such measurements across autophagy-activating conditions. These targeted interactions provide more information about autophagy than all previous datasets, producing a second-generation ontology of 220 functions. Approximately half are new, in which we confirm roles for Gyp1 at the phagophore-assembly site, Atg24 in cargo engulfment, Atg26 in cytoplasm-to-vacuole targeting, and Ssd1, Did4 and others in selective and non-selective autophagy. The procedure and autophagy hierarchy are at http://atgo.ucsd.edu/

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-37

Phylogenetic and modelling based analysis of NAD biosynthesis

Ines Heiland [1,*], Mathias Bockwoldt [1], Toni Gossmann [2], Mathias Ziegler [3]

[1]UiT Arctic University of Norway, Tromsø, Norway, [2]The University of Sheffield, Sheffield, United Kingdom, [3]University of Bergen, Bergen, Norway

Abstract: NAD is best known as cofactor for redox reactions. But in addition, it is substrate for NAD-dependent signaling reactions that are part of major regulatory pathways, including DNA-damage repair as well as histone and enzyme modifications. NAD therefore provides an important link between metabolism and cellular signalling. The most important precursor for NAD in mammals is nicotinamide (Nam), which is also the product of NAD-dependent signaling reactions. NAD homeostasis is maintained by salvage pathways that recycle Nam to NAD. Here, we present an analysis of the phylogenetic development of the pathway in eukaryotes. We reveal several co-evolving enzymes and identify important sequence variations that can be linked to known evolutionary changes in e.g. compartmentalisation of the process. Using biomathematical modelling we are able to explain the physiological relevance and the underlying molecular mechanism of the observed pathway evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES3

How universal is the minimal gene set for bacterial life?

Arpit Jain [1], Arndt von Haeseler [2], Ingo Ebersberger [1,*]

[1]Inst. of Cell Biology and Neuroscience, Goethe University Frankfurt, Frankfurt, Germany, [2]Center for Integrative Bioinformatics (CIBIV), Max F. Perutz Laboratories (MFPL), Vienna, Austria

Abstract: Orthologous sequences, the subset of homologs that separated by a speciation event, document the evolutionary history of contemporary species. In addition, they provide insights into the gene repertoires, interaction networks and metabolic capacities of primordial organisms down to *LUCA*, the last universal common ancestor of organismic life. Experimental design of artificial life challenges now the evolutionary inferences of a universal genetic repertoire common to – and necessary for – all living organisms. Only recently, 438 protein coding genes from the bacterium *Mycoplasma mycoides* were determined as the minimal gene (MG) set required for a self-replicating bacterial cell. As both the MG set and the gene set for LUCA identify a combination of genes required for cell survival and replication, it is puzzling that only about 38 % of the MG set can be traced back to *LUCA*. How can we explain this difference? Especially, why do so many genes with unknown function occur in MG that do not have a homolog in other organisms.

Here we propose, that large evolutionary distances may hamper the detection of orthologs. To substantiate this hypothesis, we developed *ProtTrace*, a tool that assesses for individual proteins over which evolutionary distances sequence similarity will suffice for an ortholog identification. Thus, we study the *Traceability* of individual genes in evolutionary time scales. Investigating the traceability of the entire yeast gene set, we find that proteins with high traceabilities in bacteria are highly enriched for catalytic functions in the cell metabolism. Yeast genes dating back to *LUCA* are almost entirely recruited from this set. On the other hand, major fractions of the regulatory network in control of gene expression have traceabilities that prevent a detection of orthologs in prokaryotes, even if they exist. These findings are mirrored in the MG set. Proteins with high traceability, many of which have functional counterparts in eukaryotes, are enriched for enzymes and ribosomal proteins. In turn, the vast majority of proteins in the MG set that appear confined to bacteria – many of which with an entirely unknown function – are of low traceablity. Extrapolating from the findings in yeast, it is now tempting to speculate that the regulatory network, which is probably relevant also for artificial life, is hidden in this fraction. The minimal gene set for bacterial life might therefore be more universal that hitherto assumed.

In summary, we show that protein traceability has a pronounced influence on both the reconstruction of the evolutionary past and on the propagation of functional annotation across contemporary species. Our results imply that cellular life is facilitated by a widely visible network of enzymatic activities. This however seems to rest on a basis of a dynamically evolving regulatory network for which most signals informing about any ancient evolutionary origin might have long been lost.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-12

Modeling genetic drift among divergent somatic lineages to estimate the effective number of stem cells in asexual planarians

Hosseinali Asgharian [1 2,*], Joseph Dunham [2], Tevfik Kitapci [2], Paul Marjoram [3], Sergey Nuzhdin [2]

[1]Biochemistry and Biophysics, University of California, San Francisco, San Francisco, [2]Biological Sciences, [3]Preventive Medicine, University of Southern California, Los Angeles, United States

Abstract: Planarians have regained attention in recent years owing to their extraordinary capacity for reconstructing whole bodies from small tissue fragments - promising to be extremely informative towards the efforts in regenerative medicine. It is estimated that stem cells comprise about 30% of their body but details of the regeneration process are largely unknown. For example, it is not clear if all stem cells or only a fraction of them close to the wound site participate actively in each round of regeneration; or whether different specialized stem cell types replenish specific tissues and organs. Due to unavailability of transgenes for these species and lack of a high quality reference genome, many routine molecular and cell biology techniques cannot be applied to this system yet. We modeled each cell as a separate individual and the body of a worm as a population of cells. We tried to estimate the effective number of stem cells based on the temporal variance of allele frequencies across 16 generations sampled every other generations. The inferred number of effective stem cell based on somatic genetic drift falls considerably below those estimated by microscopic methods. Additionally, the hypothesis of trans-generational lineage specialization of stem cell subtypes was investigated through a re-analysis of published single-cell RNA-seq data. These results improve our understanding of body regeneration through stem cells in planaria, and illustrate the application of population genetic theory as an effective analysis tool to study organismal phenomena through cellular data.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES5

Analysis of the replisome plasticity and its consequences for genome evolution

Marco Fumasoni [1,*], Andrew Murray

[1]Dept of Molecular and Cellular Biology, Harvard University, Cambridge, United States

Abstract: Cells duplicate their genomes using molecular machines called replisomes. Because of the fundamental reactions performed by this complex, many of its subunits are essential for viability. Complex interactions between a group of proteins are usually considered a constraint in the evolvability of a biological system, and a valid explanation of a complex's evolutionary conservation. Despite this view, the molecular structure of the replisome varies between bacteria, archaea and eukaryotes, suggesting that this essential complex acquired new features during evolution. We used the budding yeast, *Saccharomyces cerevisiae*, to investigate the replisome's plasticity towards genetic perturbations by experimentally evolving populations lacking Ctf4, which ensures physical coupling between the replisome's helicase and the primase activities. Evolved populations show a general recovery of most of the fitness defect induced by removing Ctf4. Analysis of other distinctive *ctf4D* phenotypes reveal heterogeneity in the behavior of the evolved populations, suggesting distinct adaptive strategies adopted by different lines during the experiment. Genome sequencing revealed putative adaptive mutations in several DNA metabolism pathways. Progress in dissecting the molecular basis of adaptation to the replisome perturbation, and the implications for genome evolution will be discussed.

Expanded summary*: BACKGROUND AND BROAD SIGNIFICANCE

Propagating genetic information from one generation to the next is an essential prerequisite for life and evolution. To accomplish this task, living organisms are endowed with multiple enzymes that catalyze individual reactions such as unwinding of the double helix, replication priming, and synthesis of the DNA. Interestingly, a shared common feature among living organisms is the organization and coordination of these enzymatic activities in complexes called replisomes. Due to the fundamental activities carried on by different replisome components, mutations in the individual genes are often incompatible with life, making the gene products essential for viability. The existence of complex interactions that control essential activities is generally used to explain the conservation of biological processes throughout evolution and the similar overall organization of the replisome in different life domains supports this concept . Yet, although similar in their topological organization, the number of the replisome subunits and their individual structures vary between bacteria, archaea and eukaryotes. This suggests that although it performs essential and biochemically conserved reactions, the replisome acquired several features during evolution and even diverged in different species. The plasticity of the replisome as well as the mechanisms that allow its adaptation are still not known. Our study of the evolvability of the replisome and its effects on the rest of the genome address these questions and reveal new insights into the evolution of essential functions and the mechanisms that maintain genome stability. We used the budding yeast, *Saccharomyces cerevisiae*, to investigate the replisome's plasticity towards genetic perturbations by experimentally evolving populations lacking Ctf4, which ensures physical coupling between the replisome's helicase and the primase activities. We performed whole genome sequencing of the evolved populations after 1000 generations and we identified the putative adaptive strategies cells have adopted to overcome the ancestor’s defects. Analysis of the mutations frequency in intermediate populations frozen throughout the experiment revealed the history of mutations accumulated in individual lines. Finally, reconstruction of the putative adaptive mutations in the ancestor strain allows us to experimentally measure the fitness advantage of each mutation and to reconstruct the order of molecular events that allowed cells to adapt to the absence of an important replisome subunit. DNA replication is among the most ancient of cellular processes. The enzymatic activities that promote it appeared billion years ago , making direct study of these evolutionary transitions impossible. The deletion of Ctf4 mimics an ancient condition when the individual activities that make up replication were less well coupled. Evolving these cells allowed the direct observation of the genome dynamics that are induced by the absence of this coordinating function. This approach offers a window into the evolution of the replication machinery and explores the general principles that govern the natural selection of essential processes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES4

Pleiotropic genetic effects on cell morphology are mediated by underlying relationships among single-cell morphological features

Kerry Geiler-Samerotte [1,*], Austin Taylor [2], Harris Lazaris [2], Chelsea Ramjeawan [2], Naomi Ziv [2], Annalise Paaby [3], Mark Siegal [2]

[1]Biology, Stanford University, San Francisco, [2]Center for Genomics and Systems Biology, New York University, New York, [3]Biology, Georgia Institute of Technology, Atlanta, United States

Abstract: Understanding the mapping from genotype to phenotype is a major goal of biology. Recent studies suggest that we can make progress toward this goal by understanding cellular systems. We prove this point when we screen for genes influencing single-cell morphology. We use high-throughput microscopy to quantify 200 morphological features for ~one million cells belonging to 378 yeast strains. We find many loci that each influence an astounding number of morphological traits. To understand the mechanism by which pleiotropic genes influence so many traits, we leverage a unique property of our dataset – thousands of clonal cells from within each strain – to learn about the underlying relationships among morphological features. We quantified correlation coefficients for every pair of traits, partitioning correlations that arise between strains from those that are also present across clones. Then we compared coefficients across subsets of clonal cells. We found the relationships between morphological features are very often multifaceted. For example, morphological features are related through geometric constraints (*e.g.* cell area often increases as nuclear area increases) as well as through cell division (*e.g.* nuclear area and brightness both increase during mitosis). These various relationships obscure the overall dependencies between traits across clonal cells. This results in genetic effects that appear to influence independent traits, even after principal component analysis. Through a better understanding of cell biology, we show that pleiotropy, in this system, does not often result from genes affecting independent cellular processes. Instead, pleiotropic genetic effects percolate through underlying networks of related morphological traits.

Expanded summary*: Evolutionary biologists often ask whether related traits can respond independently to selection. Similar questions now garner the attention of cancer biologists. New strategies suggest guiding cancers into evolutionary dead ends by selecting for tumor cells that resist a first drug but are susceptible to a second. But what if cancers escape this trap through genetic variants that disrupt the negative correlation between drug responses? I studied the conditionality of relationships between traits by exploiting unique features of clonal populations.

I used a novel statistical appraoch to quantify correlations between 200 single-cell morphological traits in yeast, partitioning relationships that result from distinct mechanisms by comparing correlations across:

(1) genetically distinct strains.
(2) clonal populations.
(3) subsets of each clonal population in the same phase of the cell-cycle or with other similar features.

The result of this study is a framework for quantifying correlations between traits and disentangling conditional trait-trait relationships from potentially unconditional ones.

The power of quantifying heterogeneity among clonal cells has often been over-looked. A very broad impact of this work might be a hint at a possible way to make predictions about how cells will respond to evolutionary traps by obtaining phenotypic measurements at the single-cell level. Predicting evolutionary responses is important; adaptive diseases (*e.g.* bacterial infections or cancer) underlie 30% of deaths worldwide. Understanding the physiological basis of adaptation (*e.g.*, whether genetic effects influence multiple phenotypes because those phenotypes are inextricably related) may enable better predictions and better evolutionary traps.

More generally, understanding the physiological mechanisms underlying how genetic effects influence multiple phenotypes may broadly improve our ability to map genotype to phenotype, which is a major goal of biology.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-21

A systematic survey of epistasis between gene deletions and evolved compensatory mutations

Jose Rojas Echenique [1,*], Michael Desai [1], Sergey Kryazhimskiy [2], Alex Nguyen Ba [1]

[1]Organismic and Evolutionary Biology, Harvard University, Cambridge, [2]Ecology, Behavior and Evolution, University of California San Diego, San Diego, United States

Abstract: Epistatic interactions underlie foundational problems in evolutionary biology: the role of history and chance in determining the outcomes of evolution, the relative difficulties of evolving different complex adaptations, and the evolution of sex and recombination. To characterize the general patterns of epistasis produced by adaptation, we evolved 20 replicate populations of 36 different yeast gene deletion mutants and measured the epistasis between the initial gene deletions and the mutations acquired in the course of adaptation. We found that rates of adaptation were genotype dependent and sought to explain this result in terms of the measured epistatic effects.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES6

Conserved transcription factors are required for patterning cnidocytes (a novel cell type) in the sea anemone Nematostella vectensis

Leslie S. Babonis [1,*], Mark Q. Martindale [1]

[1]Whitney Lab, University of Florida, St. Augustine, United States

Abstract: Cnidocytes, the stinging cells characteristic of cnidarians (corals, jellies, etc) are defined by their possession of an unusual organelle – the cnidocyst – making them one of the few clear examples of a truly novel cell type. Diverse in both morphology and function, cnidocysts are an important diagnostic feature of cnidarians; despite this, little is known about how cnidocyst diversity arises. Using a morpholino-mediated knockdown strategy, we characterize the molecular development of cnidocytes following their differentiation from a SoxB2-expressing neural progenitor cell lineage in the sea anemone, *Nematostella vectensis*. We show that two conserved transcription factors - PaxA, a paired box transcription factor, and Mef2, a MADS box transcription factor – are expressed in two different lineages of cnidocytes and are both downregulated in SoxB2 knockdown embryos. Knockdown of each of these transcription factors individually results in loss of cnidocytes. Interestingly, the *N. vectensis* orthologs of two transcription factors known to be critical for cnidocyte development in Hydra (prdl-b and COUP-TF) are unaffected by SoxB2 knockdown, suggesting that cnidocyte development in these two lineages of cnidarians involves distinct signaling pathways. Using a provisionally assembled cnidocyte gene regulatory network for *N. vectensis*, we explore commonalities in the developmental pathways of this unusual cell type across cnidarian lineages.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-18

Evolution under direct selection for growth rate

Chi-Chun Chen*, Michael Lynch [1]

[1]Biology, Indiana University, Bloomington, United States

Abstract:

Growth and division are defining features of cellular life and their rates are important proxies of fitness. While *Escherichia coli* was argued to approach thermodynamic limit of growth [1] with ~20 min doubling time, cells with less than 10 min doubling time were reported in recent years [2, 3], raising the question of what makes *E. coli* not replicating faster. As scaling relationships have been observed between growth rate and cell size both within [4] and across [5] species, further understanding of constraints on the evolution of growth rate could also shed light on the variation of cell size.

To investigate how growth rates may evolve, we design a modified version of chemostat whose flow rate is dynamically kept at maximally possible value as long as the cell culture would not be diluted away. We periodically collect samples from the evolved culture to analyze the changes in genome, expression profile, and proteome composition. Inserting these values into genome-scale metabolic models [6], we compare the evolved growth rates with those predicted by the models and discuss how the loss or modification of certain pathways could improve growth rate, for example, through optimization of energy efficiency [7]. Then we compete evolved strains with the ancestor under both stable and variable environments to test the idea that growth rate did not evolve to be maximized because of the trade-off between growing fast in one environment and being prepared for changes in environment [8].

References

[1] England, J. L. Statistical physics of self-replication. J Chem Phys 139, 121923 (2013).
[2] Elsgaard, L. & Prieur, D. Hydrothermal vents in Lake Tanganyika harbor spore-forming thermophiles with extremely rapid growth. Journal of Great Lakes Research 37, 203--206 (2011).
[3] Weinstock, M. T., Hesek, E. D., Wilson, C. M. & Gibson, D. G. *Vibrio natriegens* as a fast-growing host for molecular biology. Nat Meth 13, 849--851 (2016).
[4] Taheri-Araghi, S. et al. Cell-size control and homeostasis in bacteria. Curr. Biol. 25, 385--391 (2015).
[5] Kempes, C. P., Dutkiewicz, S. & Follows, M. J. Growth, metabolic partitioning, and the size of microorganisms. Proc. Natl. Acad. Sci. U.S.A. 109, 495--500 (2012).
[6] Lewis, N. E. et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol. 6, 390 (2010).
[7] Maitra, A. & Dill, K. A. Bacterial growth laws reflect the evolutionary importance of energy efficiency. Proc. Natl. Acad. Sci. U.S.A. 112, 406-411 (2015).
[8] Price, M. N., Wetmore, K. M., Deutschbauer, A. M. & Arkin, A. P. A Comparison of the Costs and Benefits of Bacterial Gene Expression. PLOS ONE 11, e0164314 (2016).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES8

Stochastic differences in gene expression levels contribute to the overall metabolism of a clonal cellular population

Christopher Morales [1], Kashyapa Bandaralage [1], Tuya Yokoyama [1], Joshua Rest [1,*]

[1]Dept. of Ecology and Evolution, Stony Brook University, Stony Brook, United States

Abstract: It is now clear that some genes exhibit a high level of stochastic variation ("noise") in expression levels among isogenic cells. This high variance may be caused by intrinsic or extrinsic factors, and has been hypothesized to contribute to bet hedging, stable mixed strategies, and divisions of labor. The goal of our work was to test which of these strategies contribute to the stochastic differences in the expression levels of proteins observed among isogenic cells for a panel of genes in *Saccharomyces cerevisiae*. We experimentally determined the relative fitnesses of cells that have different levels of protein expression due to expression noise. We observed that cells with different expression levels form different components of a structured population. Our experiments indicate that cells with outlier levels of gene expression for some noisy genes (e.g. Ara1, Glk1, Tsl1) exhibit substantially slowed growth when isolated from the rest of the population. This growth phenotype is rescued when outlier cells are in the presence of the total population, or in an environment that simulates metabolites present in the media of a normal population. These results suggest that cells with different expression levels are metabolically interacting with each other to form the total growth rate of the population, even as individual cells switch among alternative states. While we observed that expression outliers benefit metabolically from the main population, an intriguing possibility is that the outliers also contribute metabolic products to the main population and that the evolution of stochastic gene expression noise may serve to create structured populations with a division of labor among metabolic contributions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-17

Investigating viral attenuation by promoter knockout: a systems approach

Matthew Paff [1,*], Benjamin Jack [1], Dan Boutz [1], Bartram Smith [1], Claus Wilke [1], James Bull [1]

[1]The University of Texas at Austin, Austin, United States

Abstract: Live attenuated viral vaccines provide the most robust and longest lasting immune response. Yet designing them *a priori* to have reduced growth capacity and also to be robust to evolutionary reversion can be challenging. On the one hand, genome editing methods now enable us to create almost any conceivable viral genome composition. Yet understanding and predicting how engineered genomes will behave and evolve is a challenge. Here we adopt a systems approach in studying a simple attenuation design in bacteriophage T7: promoter knockout. Either or both promoters for the two most highly expressed genes were abolished. Overall fitnesses, major phenotypes and gene expression levels were measured for all initial genomes and for genomes evolved toward for fitness recovery. Initial genomes behaved broadly as expected, but the genomes showed an unexpected ability to evolve back to high fitness. Genome sequences, RNA Seq and proteomics reveal the molecular foundations of the attenuations and recoveries. Overall, the work suggests that a systems approach is ultimately yielding to understanding, if not predicting the consequences of genome editing and evolutionary recoveries of simple genomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-20

Evolutionary systems biology integration of multi-level CTMC interaction models of biochemistry and cancer cell growth using Evolvix

Jocelyn Meyer [1,*], Elaine Alarid [2], Laurence Loewe [1]

[1]Laboratory of Genetics and Wisconsin Institute for Discovery, [2]Department of Oncology, University of Wisconsin-Madison, Madison, United States

Abstract: While biochemistry evidently affects the growth rate of cells, many biochemists routinely ignore population variation, just like population geneticists usually ignore causal details of biochemistry that underpin a change in growth rate caused by a mutation. A true EvoSysBio integration requires an explicit mechanism for how molecular reaction rates affect the reproduction rates that determine the fitness of an organism. Here we simulate a very simple and completely explicit Continuous Time Markov Chain (CTMC) model of cancer cells whose growth rate is affected by the biochemical equilibrium between two molecular complexes. Approximately 70% of breast cancers are of a type that overexpress Estrogen Receptor-alpha (ER). Cell growth in this type of cancer is inhibited by hormonal therapies that antagonize ER function as a transcription factor. ER is encoded by the ESR1 gene, which itself is a target of ER-mediated transcription. When activated by estrogen, ER binds to the ESR1 promoter, repressing new synthesis of ER protein. Estrogen binding also induces pathways that lead to degradation of ER protein. This negative feedback loop is finely tuned to natural levels of estrogen and results in natural levels of growth. In breast cancer, the system is thrown off its natural course such that increased levels of ER induce levels of cell-growth that lead to cancer. Thus, both genetic changes to the ESR1 promoter, ER protein degradation, and biochemical changes in estrogen metabolism can effectively cause changes in the rate of cell growth, which can be seen as the 'fitness' of a cancer cell. Predicting cancer cell growth in this system raises a conceptual multi-level simulation problem, because the molecular aspects of this model need to compute the biochemistry in a way that influences growth rates at the cellular level (but without resetting growth at each cell division). We will present progress towards addressing this simulation challenge in a pure mass-action model implemented using the Evolvix model description language. In particular, we explore how a Lefkovitch-matrix might contribute to solving this generic modeling problem that reoccurs in many contexts. We invite discussions on how to best model such dynamic multi-level systems.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-15

Disrupted proteome in a virus attenuated by codon deoptimization

Benjamin Jack [1,*], Daniel Boutz [1], Matthew Paff [1], Bartram Smith [1], James Bull [1], Claus Wilke [1]

[1]University of Texas at Austin, Austin, United States

Abstract: The engineering of hundreds of synonymous codon changes into a viral genome appears to provide a general means of achieving attenuation. The mechanistic underpinnings of this approach remain enigmatic, however. By integrating quantitative proteomics, RNA sequencing, and computational modeling, we explore the molecular basis of attenuation in a strain of bacteriophage T7 whose major capsid gene was engineered to carry 182 suboptimal codons. As expected, there was no evident effect of the recoding on transcription. Proteomic observations revealed that translation is halved for the recoded major capsid gene, and a smaller reduction applies to a few genes downstream, potentially caused by translational coupling. Viral burst size is also approximately halved, and the fitness drop accompanying attenuation is compatible with the reduced burst size. Overall, the fitness effect and molecular basis of attenuation by codon deoptimization are compatible with a relatively simple model of reduced translation of a few genes and a consequent diminished virion assembly. This mechanism is simpler than that operating in eukaryotic viruses.

Expanded summary*: Live vaccine development depends on our ability to reliably attenuate viruses. Introducing many 'suboptimal' synonymous codon changes effectively attenuates many viruses and offers several advantages over earlier methods of attenuation. These recoded viruses invoke a strong immune response by producing proteins identical to that of the wild type. Moreover, since many individual codon changes cumulatively impair fitness, these attenuated viruses are evolutionarily stable. Yet, the mechanism by which silent codon changes attenuate viruses has proven increasingly elusive. Furthermore, it remains unclear whether a single mechanism underlies attenuation in different viral systems. Here, we extend previous work in a bacterial virus in which the encoding of rare codons in an essential, highly expressed capsid gene reduced fitness. We utilize RNA sequencing, quantitative proteomics, computational modeling, and various phenotypic measures to elucidate the viral life-history effects of codon deoptimization. Our goal is to refine an understanding of the molecular basis of attenuation.

To understand why recoding affected fitness, we investigate a previously characterized strain of bacteriophage T7 with a recoded capsid protein (gene *10* in standard T7 genomic nomenclature). The capsid protein is the most abundant protein produced during the infection cycle of bacteriophage T7. In the recoded strain, fully 182 codons had been replaced in the gene *10*, reducing fitness from ~42 to ~37 doublings per hour. This reduction corresponded to a 180-fold decrease in phage particles produced per hour. Here, we extend this work to explore the underlying molecular mechanism by which the recoding reduces fitness. Our primary result is that the protein product of gene *10* is reduced almost 50% by the end of the infection cycle, but protein abundance of genes immediately downstream of gene *10* are also depressed. The differences in protein abundance are not reflected in transcript levels, so it appears that the suppression of protein levels lies in translation. Burst size is also halved in the recoded strain. The evidence thus supports a simple interpretation of the fitness impact of recoding the major capsid gene: (i) capsid protein is expressed at a reduced level, as are a few downstream genes, and (ii) burst size is correspondingly reduced approximately 50% with no change in lysis time, compatible with the observed reduction in total fitness.

That recoding gene *10* also affected expression of downstream genes is an unexpected result. We propose that translational coupling may explain why expression of genes downstream of *10* is suppressed by the recoding. This coupling is plausible because T7 produces many polycistronic transcripts. Moreover, both biophysical and mathematical models of translational coupling produce predictions compatible with our observations. Although our evidence is indirect, this study may provide the first indication that translational effects of the recoding extend beyond the recoded genes.

The approach developed in this study, and the data generated, will be useful for formulating computational models that scale from the molecular level to that of viral fitness. At the molecular level, we conducted RNA-sequencing and quantitative proteomics analyses from phage-infected *E.* coli sampled at 1, 5, and 9 minutes after infection. Our detailed time-course data could form the basis of *in*

silico fitness predictions of codon deoptimization. Ultimately, we envision a future in which an understanding of viral life history at the molecular level enables facile engineering of arbitrary fitness and alternative vaccine designs.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-19

A population phylogeny approach to understanding mitochondrial heteroplasmy

Peter Wilton [1,*], Thorfinn Korneliussen [1], Marcia Su [2], Arslan Zaidi [2], Kateryna Makova [2], Rasmus Nielsen [1]

[1]Integrative Biology, UC Berkeley, Berkeley, CA, [2]Department of Biology, Penn State University, University Park, PA, United States

Abstract: Genetic variation of mitochondria within the body, termed heteroplasmy, is the major cause of mitochondrial disease and has roles in aging and tumorigenesis. Recently, heteroplasmy has also been shown to be a part of healthy human biology. Heteroplasmy may arise de novo in an individual, or it may be inherited from the mother via the mitochondria contained in the oocyte. Here, we develop a novel theoretical approach to understanding heteroplasmy by modeling the mitochondria of several tissues sampled from a family as populations related by an ontogenetic phylogeny reflecting human development and reproduction. In this framework, we calculate a likelihood for the observed heteroplasmy frequencies, accounting for genetic drift, mutation, and natural selection, and perform inference using Bayesian MCMC. We apply this inference procedure to an in-house dataset of mitochondrial sequences sampled from multiple tissues in >100 human families, as well as to multiple publicly available datasets. By inferring how different population-genetic forces act along each branch of the ontogenetic phylogeny, we learn about the processes shaping genetic variation in mitochondria within an individual and between related individuals. We find that genetic drift, mutation, and natural selection each have a role in explaining mitochondrial genetic variation at different stages of human life, thus elucidating previously unknown aspects of cellular differentiation. This work both refines previous estimates of the germline mitochondrial bottleneck, an important determinant of mitochondrial disease inheritance, and provides a new, quantitative understanding of the proliferation of mitochondria in the body and the differentiation of cell types during ontogenesis.

Expanded summary*: Background. Ever since the first assays of genetic diversity, it has generally been understood that organisms possess just a single mitochondrial genome, which they inherit from their mother. Gradually over the last few decades, it has become clearer that genetic variation amongst the mitochondria in the body often exists, and that such variation can be harmful to the health of the organism. Mitochondrial disease is now known to be caused in large part by this intra-organism mitochondrial genetic variation, termed heteroplasmy, and it has been found that mitochondrial diseases present differently depending on the frequency of the disease-associated heteroplasmic allele in one tissue versus another [1]. In contrast to many other diseases with known genetic causes, patterns of inheritance of mitochondrial disease are difficult to characterize owing to the non-Mendelian mechanism of inheritance of mitochondria [1].

In order to improve our understanding of how mitochondrial disease is inherited, it is necessary to have a more complete understanding of the transmission of mitochondria from mother to offspring and the proliferation of mitochondrial in the developing human organism. Direct observation of these cellular processes may be feasible in mice and other mammalian model organisms, but in humans an indirect approach must be taken. We have developed a novel approach to understanding the processes generating and maintaining heteroplasmy by viewing the mitochondria in different tissues as populations related by an ontogenetic phylogeny. With this perspective, it is possible to use the tools of population genetics to interpret patterns of genetic variation amongst different mitochondrial tissues and learn about the cellular processes that shape this genetic variation during development and during adult life.

Data. We have collected a dataset of mitochondrial sequences sampled from two tissues of different germ layer origins (buccal mucosa and blood) in each individual of more than >100 healthy human families. In order to better discover the likely abundant low-frequency heteroplasmies, each sample was sequenced to high depth (~5000x). Because the modeling of low-frequency mutations in pooled sequencing data (as we have, here) requires careful modeling of sequencing error, we have also generated replicate test sequence datasets that allowed for position-specific sequencing error profiles and a

probabilistic model for PCR duplication errors. This error modeling is directly incorporated into our inference procedure (see below).

In addition to this original dataset, we also analyze a previously published dataset of mitochondrial sequences from 12 tissues in 152 human cadavers, sequenced to high coverage [2]. This dataset does not contain related individuals, so nothing can be learned about the inheritance of heteroplasmy, but a population-phylogenetic analysis of this dataset brings new understanding of the dyanmics of mitochondrial proliferation during development and adult life in a variety of somatic tissues.

Model. We have developed a full probability model of the transmission of mitochondria between generations and their proliferation within a developing body. For each family from which we sample mitochondria, there is a ontogenetic phylogeny relating the sampled tissues. Along each branch of this phylogeny, the frequency of a mitochondrial mutation may change during the ontogenetic process corresponding to the branch. Thus by inferring the action of genetic drift, mutation, and natural selection along each branch in this ontogenetic phylogeny, we can pinpoint the timing of these different forces acting on mitochondrial genetic variation within the body and between mother and offspring.

We use the classic pruning algorithm [3] to calculate the likelihood of observed heteroplasmy frequencies at the leaves of the phylogeny (i.e., in the sampled adult tissues). Allele frequency transition distributions are calculated using a novel approximation to the Wright-Fisher diffusion: transition distributions are precalculated at a variety of time-points, mutation rates, and selection coefficients, and any desired distribution is calculated by linearly interpolating between distributions at precalculated values. The efficiency of this transition probability calculation allows us to perform inference using Bayesian MCMC, providing a straightforward way of quantifying uncertainty in our estimates. Uncertainty in heteroplasmy frequencies is incorporated into our inference procedure by integrating over allele frequency likelihoods calibrated using our sequencing error dataset (see above). Our method is flexible with regard to the family structure or relationship amongst the sampled tissues and thus is easily applied to a variety of datasets.

Impacts. A population-phylogenetic understanding of mitochondrial heteroplasmy has a number of biomedical implications. It is not currently possible to accurately predict the probability of inheritance of mitochondrial disease based on observations of heteroplasmy in somatic tissues at a disease-associated allele in the mother. This work represents a crucial step towards enabling such predictions. Additionally, our inferences of the rate of accumulation of mutations in mitochondria in the adult human body have implications in the study of aging and carcinogenesis, as heteroplasmy plays a role in these processes [4]. Finally, this framework provides a novel approach to studying the cell lineage of somatic stem cells in humans, as those tissues that are replenished by the same population of somatic stem cells will share mitochondrial variants, a signature that is directly assessed in this framework. Mitochondrial heteroplasmy has been used to study stem cell lineage at a fine scale in epithelial tissues [5], but a more general population phylogeny approach across a range of tissues, as provided by our analysis of the data from 12 human tissues, has yet to be conducted.

References

1. Schon, E. *et al*. *Nat Rev Genet* **13,** 878–890 (2012).
2. Li, M., *et al*. *PNAS* **112,** 2491–2496 (2015).
3. Felsenstein, J. *Syst Biol* **22,** 240–249 (1973).
4. Stewart, J. B. & Chinnery, P. F. *Nat Rev Genet* **16,** 530–542 (2015).
5. Walther, V. & Alison, M. R. *WIREs Dev Biol* **5,** 103–117 (2016).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-16

Use of network-based approaches to shed light on patterns of prokaryotic gene transfer to eukaryotes

Cedric Bicep*, Tom Williams, Robert Hirt, Martin Embley

Abstract: Lateral gene transfer (LGT) is known to be important for the evolution of prokaryotes but the importance of LGT in eukaryotes is still controversial. It is now thought that eukaryotes emerged from an endosymbiosis between an alphaproteobacterium and an archaeal host. Therefore, the existence of many prokaryotic-like genes in eukaryotes may be the legacy of eukaryogenesis. A number of reports in the literature have also suggested more recent prokaryote-to-eukaryote LGT outside of endosymbiosis, but the relative importance of this more recent LGT is still controversial. In the present study we have investigated the extent of recent LGT for a representative sample of eukaryotic genomes using a common set of methods. We used network-based methods and graph theory algorithms to analyze 34 eukaryote genomes covering all the supergroups of the eukaryotes, and also included prokaryotic reference genomes from 821 bacteria and 21 Archaea. To identify patterns likely to represent prokaryote-to-eukaryote LGT, we combined the topological properties of eukaryotic sequences in a similarity network with the taxonomical information associated with sequences, and identified cases where eukaryotic sequences show an atypical distribution. Our analyses suggest that there has been a continuous flow of prokaryotic genes involved in the replacement of previously existing eukaryotic genes by a prokaryotic homologue. The LGTs we detect appear to differentially affect all eukaryotes and are involved mainly in eukaryotic metabolism affecting different metabolic pathways to different degrees.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES9

Gene duplication imparts robustness and fragility to protein interaction networks

Christian Landry*, Guillaume Diss, Isabelle Gagnon-Arsenault, Anne-Marie Dion-Coté, Caroline Berger, Hélène Vignaud

Abstract: The maintenance of duplicated genes is thought to protect cells from genetic perturbations but the molecular basis of this robustness is largely unknown. By measuring the interaction of yeast proteins with their partners in wild-type cells and in cells lacking a paralog, we found that 22 out of 56 paralog pairs compensate for the lost interactions. An equivalent number of pairs exhibit the opposite behavior and require each other's presence for maintaining their interactions. These dependent paralogs generally interact physically, regulate each other's abundance and derive from ancestral self-interacting proteins. This reveals that gene duplication may actually increase mutational fragility instead of robustness in a large number of cases.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-26

Evolution of lipid concentrations in mammalian tissues

Ekaterina Khrameeva [1 2,*], Ilia Kurochkin [1], Philipp Khaitovich [1 3 4]

[1]Skolkovo Institute of Science and Technology, [2]Institute for Information Transmission Problems, Moscow, Russian Federation, [3]CAS-MPG Partner Institute for Computational Biology, Shanghai, China, [4] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Abstract: Lipids are essential structural and functional components of cells. However, little is known about lipidome composition in different tissues, as well as about its evolutionary dynamics across mammalian species. Here we report a large-scale analysis of the evolution of lipid composition in six tissues across 35 species that represent three major mammalian phylogenetic clades (primates, rodents and bats). We show that lipids with human-specific concentrations in cortex are linked to enzymes with human-specific gene expression changes and with a significant excess of Neanderthal ancestry, suggesting a possible role of these lipids in evolution of brain functioning on a human lineage. Moreover, lipids with concentration differences between species following phylogenetic distances demonstrate greater average divergence among species, apparently evolving faster due to directional selection.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-24

Cancer spatial evolution in a changing microenvironment

Xiaowei Jiang [1,*], Ian Tomlinson [1]

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

Abstract: Cancer development as an ecological and evolutionary process is poorly understood, which includes early cancer evolution, malignancy and metastasis. It was hypothesised that tumour microenvironment (TME) plays a critical role in this process. Unfortunately in most cancer modelling studies the TME is ignored or considered static and different cancers are often studied in isolation. There is a lack of a general theory of cancer adaptive evolution (CAE). Here I establish a genetic and phenotypic model of cancer three-dimensional (3D) spatial evolution in a changing TME. With 3D individual-based simulations I show how cancer cells adapt to diverse changing TME conditions and selection intensities. I am able to capture key histological characteristics of various cancer forms including complex dynamics of spatial-temporal heterogeneity of clonal fitness and clonal mixing, ball-like and non-ball-like clonal structures. Moreover, I identify key evolutionary and phylogenetic patterns of CAE under various combinations of phenotypic, genetic, population genetic and changing TME conditions. I show classical drivers, mini drivers, Darwinian and neutral/nearly neutral evolution and cost of complexity. I demonstrate the importance of ecology in CAE. I show that there are fundamental differences in the mode of CAE when the TME is changing, which is the limiting factor of CAE. Finally, I discuss important implications for cancer evolution theories and cancer personalised medicine.

Expanded summary*: In this work I rethink cancer development from the ground by developing a three-dimensional (3D) spatial model of CAE in a changing TME. I use a formal adaptation theory—the Fisher's geometric model, which originated from the great R.A. Fisher. Fisher invented his phenotypic geometric model in his 1930 original work "The genetic theory of natural selection" to understand the nature of adaptation in Darwin's theory of evolution by Natural selection. In almost a century this model has been extended into a general form, which includes parameters such as biological complexity, epistasis, robustness and changing environment. This framework now can be used to explore many fundamental evolutionary processes and theories and can even incorporate seemly competing theories, such as the neutral/nearly neutral and selection theories of molecular evolution. However, it has never been applied to understanding cancer development as an ecological and evolutionary process. In my results the whole adaptive process can be visualised and recorded in real-time, which includes a front-end 3D graphic interface and background spatial adaptation model based on Fisher's framework. To my surprise my model is able to capture many important evolutionary patterns, such as complex dynamics of spatial-temporal heterogeneity of clonal fitness and clonal mixing, selection driven and neutral/nearly-neutral evolution, classical drivers and mini drivers. I predict the TME is the limiting factor for CAE although mutation, mitotic recombination and chromosome instability can facilitate adaptation. Using my model I can also demonstrate several fundamental theories in cancer, e.g., cancer cell origin, namely, cancer stem cell and non-cancer stem cell. For latter our lab has provided experimental evidence published in Nat. Med. Finally, I show that there is a cost of complexity associated with CAE, which is the rate of adaptation (as well as the mean fitness of the population) decreases and the mean selection coefficient increases during adaptation when the cancer cell has an increased number of traits and the TME changes. This is consistent with a cancer reverse evolution theory that cancer removes non-essential multicellular life traits by random mutations and only focuses on few important traits for unicellular life adaptation. Ideally these few traits may be cancer's Achilles' heel. I show that there are fundamental differences regarding cancer spatial-temporal evolution when the TME changes, which is not possible in any previous cancer models. I think my work is of high importance and of general interest.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-27

Dating the Evolution of Molecular Function

Alan Beavan*, David Robertson, Mark Reardon

Abstract: Proteins function by interacting in subsystems such as complexes and pathways. Despite this, when evolutionary biologists analyse the evolution of genes, they tend to ignore these functional associations, instead examining the their evolution in isolation. Here I present a novel approach to understanding the evolution of biological function. I have assigned genes to functional groups according to the interactions they are involved in then explored the evolution of these groups. Using a new algorithm, unique in its ability to find the age of duplicate genes, dates of origin have been assigned to all human genes. Using these, I analysed the differences between young and old functions. Ancient functions are enriched in metabolism, cellular component biogenesis, and multi-organism processes. Younger functions include those involved in signalling, immunity and regulation, suggesting some aspects of these processes evolved more recently. A previous analysis found a correlation between whole genome duplicated genes and dominant mutations in disease. Another has found that genes ancient in origin are more likely to be involved in genetic disease. I will present an analysis of whether these correlations are observed when analysis is restricted to individual functions, in a more systems oriented approach. I will also present visualisations of ancestral protein interaction and metabolic networks, which will elaborate on changes that occurred in human evolution at a systems level.

Statement: I am studying for an undergraduate masters (MSci) genetics at the University of Manchester. This year I am conducting a six-month research project, under the supervision of Professor David Robertson. During this project I have extensively expanded my knowledge and experience of using computational techniques, including python and R. This has helped shape my interest, and allowed me to focus on molecular evolution when considering PhDs starting next year, for which I am at the interview stage. My project has already produced significant findings and there is still much time left. This enhances my desire to explain my analyses to others at the biggest conference in molecular and evolutionary biology. I am convinced that my future is in academia, specifically in the field of genome evolution. For this reason I would like to receive the advice of some of the scientists that will be attending. This also means I will be going to many more conferences in the future, so any experience in attending the conference will be useful. I am excited to attend some of the symposia. One that is of particular interest to me is “Calibrating the history of life”. Davide Pisani is somebody who I have applied for a PhD with and the subject matter would be of great use to me. Also, I would be really excited to hear Michael Lynch talk on the evolutionary systems biology of cells. As a scientist, he has contributed so much to our understanding of evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES11

Unique distributions of the Sa and Dam1 complexes support functional analogy and suggest multiple parallel displacements of Ska by Dam1

Jolien van Hooff [1 2 3,*], Berend Snel [2], Geert Kops [1 3 4]

[1]Hubrecht Institute, [2]Biology, Utrecht University, [3]Molecular Cancer Research, [4]Cancer Genomics Netherlands, UMC Utrecht, Utrecht, Netherlands

Abstract: Faithful chromosome segregation relies on kinetochores, the large protein complexes that connect chromatin to spindle microtubules. Although human and yeast kinetochores are largely homologous, they track microtubules with the unrelated protein complexes Ska (Ska-C, human) and Dam1 (Dam1-C, yeast). The analogy of these complexes poses questions about their evolution. We here uncovered that Dam1-C and Ska-C are both widespread among eukaryotes, but in a strikingly inverse manner, supporting their functional analogy. Within the complexes, all Ska-C and various Dam1-C subunits are ancient paralogs, hence gene duplication shaped these complexes. We inferred that Ska-C was present in the last eukaryotic common ancestor, that subsequently Dam1-C displaced Ska-C in an early fungus and got horizontally transferred to diverse non-fungal lineages, displacing Ska-C in these lineages too. Our study suggests an exceptional case of eukaryotic horizontal gene transfer of components of a core cellular machinery.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES1

Mutation, Drift, and the Origin of Cellular Features

Michael Lynch*

Abstract: Although natural selection may be the most powerful force in the biological world, it is not all powerful. Consequently, many aspects of evolution at the molecular level can only be explained by the inability of natural selection to operate. This general principle explains a lot about the evolution of genome architecture, but surprisingly, it also appears to extend to multiple higher-level features of cells. Several factors conspire to make this so, including multiple degrees of freedom for constructing cellular features and the inability of selection to promote the efficiency of cellular features beyond the barrier imposed by random genetic drift. The drift-barrier hypothesis has general implications for all aspects of cellular evolution, including the replication and transcriptional fidelity, the performance of enzymes, the stability and multimeric nature of proteins, and the refinement of transcription-factor binding sites. An open question considers how often natural selection encounters a barrier imposed by biophysical laws prior to meeting the drift barrier. Finally, the drift-barrier hypothesis suggests obvious problems with the popular concept of biological robustness. Although selection can promote new attributes that substantially improve performance, the improvement will often be transient, with the now more complex system then drifting back to its prior level, yielding a system that looks more robust, while also being more vulnerable to breakdown.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES14

A Cellular Systems Approach to Origins of the Metazoa: Functional Constraints and Sequence Conservation in Coevolution of the Core Adhesome

Richard McCann*

Abstract: The independent emergence of multicellularity in multiple lineages of the eukaryotic radiation was a major transition in the evolution in each of these branches of life. The genome sequences of model unikonts (Amorphea) representative of the antecedents of the metazoan radiation have revealed that a group of conserved, interacting proteins comprises a core adhesome that was essential for the development of cell adhesion assemblies required for animal multicellularity. These include talin, vinculin, α-actinin, paxillin, and integrin, and these proteins represent the foundation upon which the more complex consensus adhesome of animals was built. As the complexity of these multicomponent assemblies increased during the metazoan radiation, the structural constraints on the individual protein components of the core/consensus adhesome also increased. This should be reflected in greater evolutionary conservation of these proteins from later diverging taxa, compared with adhesome components from model unicellular unikonts with simpler consensus adhesomes, and also to "housekeeping" proteins that are not part of complex assemblies. Comparison of the molecular clocks of adhesome components across 1.8 billion years of evolution is consistent with this hypothesis. Sequence similarity of core adhesome protein orthologs correlates positively with adhesome complexity and negatively with divergence time and less adhesome complexity. The results of this systems approach to the evolution of cellular function will provide a foundation for the study of adhesome protein coevolution during the inferred prehistory of animal multicellularity and throughout the metazoan radiation, beginning with the ancestors of extant ctenophores. The systems approach described here also will be applicable to studies of neo- and sub-functionalization of essential adhesome proteins such as talin and integrin, as well as protein coevolution in other unikont/Amorphea multicomponent assemblies critical to cellular function, and similarly to other multicomponent protein assemblies in other domains of life. Given the importance of the dysregulation of multicellularity in disease, this cellular systems approach to the fundamentals of adhesome function will have broader applications in human biology and medicine.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-23

Large-scale discovery of conserved plant protein complexes

Claire McWhite [1,*], Ophelia Papoulas [1], Kevin Drew [1], Claire Palmer [1], Cuihong Wan [2], Daniel Boutz [1], Hong Qiao [1], Karen Browning [1], Andrew Emili [2], Pamela Ronald [3], Edward Marcotte [1]

[1]The University of Texas at Austin, Austin, United States, [2]University of Toronto, Toronto, Canada, [3]University of California, Davis, Davis, United States

Abstract: Protein physical interaction networks are a roadmap to protein characterization; however, current plant protein complex maps only cover a small neighborhood of biological function and plant diversity. Here, we present conserved plant protein complexes determined from tag-less proteomic analyses of rice, wheat, Arabidopsis, Selaginella, Ceratopteris, broccoli, and the single-celled algae Chlamydomonas. Our method both recovers known complexes and identifies novel complexes that have been conserved in plants since the green algal ancestor. To define physical protein-protein interactions, specimens are first fractionated along a biochemical gradient and protein complex maps built from proteins with consistently co-elute in biochemical fractions. We find particularly interesting complexes composed of metabolic proteins, indicating a systematic trend in plants of physically organizing pathway enzymes. To allow direct comparison of proteomics experiments from an arbitrary number of species, of any ploidy, we developed a method to sort proteomes into orthologous groups and modified the mass spectrometry peptide lookup to identify orthologous groups instead of individual proteins. In addition to discovery of complexes, our dataset of over 1000 mass spectrometry analyses of diverse plants allows profiling of protein and protein complex evolution over plant evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES10

High-resolution fitness mapping of a metabolic pathway

Harry Kemble [1,*], Catherine Eisenhauer [1], Audrey Chapron [1], Jeremie Chatel [1], Melanie Magnan [1], Hervé Le Nagard [1], Philippe Nghe [2], Olivier Tenaillon [1]

[1]IAME, UMR1137 , INSERM, [2]CNRS UMR 8231 , ESPCI, Paris, France

Abstract: The fitness landscape is one of Evolution's key concepts. Metabolic pathways provide highly relevant systems-level models to examine such landscapes. Further, Metabolic Control Theory (MCT) provides a mechanistically derived framework for predicting and interpreting the properties of these landscapes. MCT has proven capable of predicting microbial fitness from enzyme activity, but only in the simplest case of directional selection, where fitness is directly proportional to pathway flux. A more relaxed assumption is that there is a pathway flux or steady-state metabolite concentration which is *optimal* for fitness, which is particularly important when buildup of a pathway metabolite is toxic. We sought to characterise such a landscape empirically, choosing as a model the first two steps of the *E. coli* L-arabinose catabolic pathway, whose product (L-ribulose-5-phosphate), like many phosphorylated intermediates, is toxic at high levels. Using a combination of chemically controllable promoters and high-throughput promoter mutagenesis coupled with NGS-based competition assays and standard growth measurements, we have mapped regions of this fitness landscape at unprecedented resolution and have uncovered striking, asymmetric patterns of epistasis which are in line with MCT predictions. The resulting landscape differs markedly from classical abstract phenotypic models of adaptation, and as such bears important implications for several open questions in adaptive landscape theory.

Expanded summary*: The fitness landscape is one of Evolution's key concepts. Metabolic pathways provide highly relevant systems-level models to examine such landscapes. Further, Metabolic Control Theory (MCT) provides a mechanistically derived framework for predicting and interpreting the properties of these landscapes.

MCT has proven capable of predicting microbial fitness from enzyme activity, but only with a small number of data points under consideration (<50), and only in the simplest scenario of directional selection, where fitness is assumed to be directly proportional to pathway flux. A more relaxed assumption is that there is a pathway flux or steady-state metabolite concentration which is *optimal* for fitness, which is particularly important when buildup of a pathway metabolite is toxic. We sought to characterise such a landscape empirically, choosing as a model the first two steps of the *E. coli* L-arabinose catabolic pathway, whose product (L-ribulose-5-phosphate), like many phosphorylated intermediates, is toxic at high levels.

The relevant two genes (*araA* and *araB*) were placed under the control of independently inducible promoters. A library of promoter mutants based on this plasmid was constructed, consisting of all possible single-bp substitutions over the 2 RNA-pol-binding hexamers in either one or both promoters. Genotype-fitness maps were obtained from the library by NGS-based pooled competition experiments. By performing these experiments under different combinations of inducer concentrations, we were able to map promoter genotype to fitness in different regions of the fitness landscape, at unprecedented resolution. A global, coarse-grained phenotype-fitness mapping was also carried out by directly measuring growth under various inducer concentration combinations.

Our results reveal striking, asymmetric patterns of epistasis between the two genes, which are in line with predictions derived from a combination of MCT and a Fisher's Geometric Model-type stabilising selection assumption, as developed by Szathmary (*Genetics* 1993). The present study demonstrates how bottom-up and top-down approaches to studying adaptive landscapes can be combined, avoiding the need for painstaking characterisation of biochemical parameters while maintaining a grounding in biological reality. Further, the interesting inferred shape of the landscape studied here has implications for open evolutionary questions such as Orr's cost of complexity and Hartl *et al.*'s selective neutrality.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES7

Comparative single-cell transcriptomics in sponges sheds light on the origins of animal cell type diversity

Jacob Musser [1,*], Detlev Arendt [1], Kaia Achim [1], Michael Nickel, Warren Francis [2]

[1]European Molecular Biology Laboratory, Heidelberg, [2]Ludwig-Maximilians-University of Munich, Munich, Germany

Abstract: A key transition in early animal evolution was the origin of specialized cell types, including the evolution of distinct types of contractile and nervous system cells. Sponges diverged early in animal evolution and have relatively few types of cells, which lack clear homology to bilaterian cell types. Despite this, sponge genomes contain homologs of genes important in muscles, neurons, and other specialized cell types. Here we present a comparative single cell-transcriptomic study in sponges that sheds light on cell type diversification in early animals. We used an unbiased approach to characterize cell types in two species of demosponge, first disassociating entire mature individuals, and then conducting single-cell RNAseq on hundreds of cells. We identify a diverse array of cell types in sponges. This includes epithelial-like cells expressing cell adhesion and signal transduction genes, collagenous cells, and ciliated cells enriched for genes implicated in sensory perception. We find that homologs of neuronal genes exhibit distinct patterns of expression. For instance, genes involved in exocytosis are broadly expressed across different sponge cell types. We also uncover genes implicated in neurotransmitter synthesis and reception that are specific to different sponge cell types. Our results suggest that although sponges lack a morphologically-distinct nervous system, they differentially deploy genes important in sensing the environment and intercellular communication.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-30

Regulatory Evolution of Cell Type Specification in Echinoderms

Gregory Cary [1,*], Brenna McCauley, Alys Cheatle Jarvela, Rene Francolini, Alyssa Lawler, Katherine Huang, Miwa Shirai, Veronica Hinman [1]

[1]Carnegie Mellon University, Pittsburgh, United States

Abstract: Echinoderm embryos present an exceptional system through which to understand the mechanisms of gene regulatory network (GRN) evolution. This is due to intricate knowledge of the sea urchin developmental GRN amassed in recent decades as well as the relative ease of interrogating both alterations to cis regulatory sequences and transcription factor function in these species. We will present recent data that leverages knowledge of the sea urchin GRN to understand the evolution of mesodermally derived cell types in sea urchin, sea star, and sea cucumber embryos. Sea star and sea cucumber embryos develop with differing proportions of blastocoelar mesenchymal cells and coelomic epithelial cells. The regulatory logic controlling the specification of these cell types involves a distinction in the role of delta/notch signaling, which further contrasts with the role of delta/notch signaling in the sea urchin endomesoderm GRN. Finally, we consider the evolution of one regulatory node in particular, the transcription factor Tbrain, and the functional consequences of an observed change in binding site affinity between the sea urchin and sea star orthologs. Using whole-genome assays to identify targets of the orthologous transcription factors in each species, we show that the proteins share very few common targets. The suggests that the network surrounding this node in each species has been dramatically re-wired since they last shared a common ancestor. Taken together, these data underscore the extensive changes to the sub-network controlling mesodermal cell type specification and highlight the need for systems-level approaches to untangle problems of GRN evolution.

Expanded summary*: The detailed knowledge of sea urchin development embedded in the gene regulatory network (GRN) provides an unprecedented platform to understand the mechanisms of network evolution. The Hinman lab has routinely used species from other classes the in the phylum Echinodermata, for example sea stars and sea cucumbers, to interrogate the changes in GRN topology and function. The approach involves first identifying differences in network topology between species, then further understanding how alterations to individual nodes (i.e. transcription factors) or edges (i.e. cis regulatory sequences) have influenced the observed changes.

In this work, we focus on the sub-network involved in the specification of mesodermal cell types in developing echinoderms. We describe how differential utilization of the delta/notch signaling pathway leads to differing proportions of two different types of mesodermal cells, blastocoelar mesenchymal cells and coelomic epithelial cells. The network inputs into this delta/notch signal in the sea star are defined, and then contrasted with the corresponding sub-network from the sea urchin GRN.

Recent work from the lab has demonstrated that a single node in this sub-network, the transcription factor Tbrain, has evolved a changed preference for DNA binding motif between the sea urchin and sea star orthologs. However, the functional implications for this under-appreciated mechanism of network evolution have not been determined. We utilize genome-wide datasets to define the targets of the ortholog from each species. We find that the two proteins share very few common target genes and functional enrichments of target genes points to a much diverged function in these two species. While we have not fully answered the

question of how evolved binding preference changes affect function, it is clear that these proteins have dramatically different roles in their respective GRNs.

These studies demonstrate the power and utility of echinoderm embryos as models to investigate the mechanisms of network evolution. The inherent complexity of network evolution is rapidly revealed and demands the implementation of systems-level experiments. Recent improvements to echinoderm genome assembly and annotation permit the application of various genome-wide techniques in these species and this is the first study to make use of these methods for comparative studies in these taxa.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-32

Extensive divergence of eukaryotic transcription factor sequence specificity

Samuel Lambert [1,*], Ally Yang [2], Debashish Ray [2], Gwendolyn Cowley [3], Mark Caddick [3], Quaid Morris [1 2 4 5], Matthew Weirauch [6 7], Timothy Hughes [1 2]

[1]Department of Molecular Genetics, [2]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada, [3]Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom, [4]Department of Computer Science, [5]Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada, [6]Center for Autoimmune Genomics and Etiology (CAGE) and Divisions of Biomedical Informatics and Developmental Biology, Cincinnati Children's Hospital Medical Center, [7]Department of Pediatrics, University of Cincinnati, Cincinnati, United States

Abstract: Transcription factor (TF) DNA-binding specificities (motifs) are often conserved between species, and many TFs have conserved functions over long evolutionary distances. As a consequence, changes in gene expression are largely attributed to evolution of *cis*-regulatory DNA sequences rather than changes to TF coding sequences. However, the number and diversity of TFs vary greatly between species, and there are numerous examples of TF diversification in DNA-binding and protein-interaction specificities. To our knowledge, TF motif evolution has previously not been explored comprehensively, and its prevalence is therefore unknown. Here, we estimate the degree of motif conservation between eukaryotic species at varying evolutionary distances. We developed an improved method to classify pairs of TFs as having similar or dissimilar motifs on the basis of their DNA-binding domain (DBD) sequences using an expanded library of DNA-binding data. We find that motif conservation is the exception between major lineages (*e.g.* metazoans, plants, fungi) due to the prevalence of clade-specific DBDs. Surprisingly, within major lineages it is typical that only half of the TF motifs are conserved since the last common ancestor, indicating that diversification has occurred in multiple TF families. Even relatively small evolutionary distances (*e.g.* human to mouse) are associated with marked alterations in TF motifs, mainly in gene families with many duplications. Our general method can be adapted to study RNA-binding proteins, enabling the comparison of the rates of divergence in regulators at multiple levels of gene expression. Overall, our study reveals that evolution of *trans* regulators is more prevalent than is generally appreciated.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-31

Lower Transcriptional Error Rates are Selected for in Highly Expressed Genes

Kendra Meer*, Paul Nelson [1], Kun Xiong, Joanna Masel [1]

[1]Ecology and Evolutionary Biology, University of Arizona, Tucson, United States

Abstract: Traverse & Ochman (2016) used Circ-Seq to measure the rate of transcription errors in *E. coli* as 10^{-4}-10^{-5}, depending on the type of substitution error, with C→U errors occurring 10 times more often than other types of error. We questioned whether this level of C→U errors was due to in vivo deamination or just an artifact of the experimental preparation. If errors occur post-transcriptionally in vivo via attack-by-water on the mRNA, then their error rates should be proportional to mRNA lifetime. Plotting the error rate per site as a function of lifetime revealed an upward trend in C→U errors, confirming the in vivo deamination hypothesis. However, graphs of all non-C→U substitutions types, which we expected to show no trend, had negative slopes. As a gene's mRNA decay rate is inversely proportional to its overall expression into protein, we posited that this observed negative trend was due to expression level. Our results could be explained if selection more strongly avoids transcription errors for highly abundant proteins. Using PaxDB estimates of protein abundance, we used linear models to confirm that results for non-C→U substitution types were driven by abundance in general rather than mRNA lifetime in particular. This further supports the hypothesis that transcriptional error rates are subject to more effective selection in genes for which a high rate of errors would be most detrimental.

Statement: I am in my junior year of undergraduate study at the University of Arizona, studying Molecular and Cellular Biology with a concentration in Information Science. My biology and chemistry courses often fill with students who strive for medical school, but my foundation in data science, and passion for the fields of genetics and genomics, direct my path towards a research career. Despite being one of few females in my more demanding programming and systems courses, I have never felt deterred from including it in my track of study. While often neglected by undergraduate biology students, I believe that a knowledge of computer systems is as essential to research as a knowledge of the processes behind the biology. Peers, and occasionally even professors, in both my biology and computer sciences courses are often confused by my desire to pursue Bioinformatics as a career, usually due to a lack of understanding of what it really is. This reaction furthers my motivation to seek opportunities, such as this conference, to deepen my understanding of what this realm of discovery is like. This opportunity to attend my first conference would provide me with a professionally and academically rewarding experience to present my research findings for assessment and criticism from other scientists and to learn from other presenters.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-29

Resolving differences on the chromosomal distributions of Drosophila new genes

Júlia Raíces [1], Paulo Otto [1], Maria Vibranovski [1,*]

[1]Genetics and Evolutionary Biology, University of Sao Paulo, Sao Paulo, Brazil

Abstract: New genes are male biasedly expressed due to their higher expression in testis than in ovaries [1]. Early studies of retrogenes and male biased genes in *Drosophila* revealed their preferential location on the autosomes in comparison to the X chromosome [1,2]. Several hypotheses such as meiotic sex chromosome inactivation (MSCI) and sexual antagonism have been proposed, and successfully tested, to explain the X chromosome demasculinization [1,3,4]. Later, studies in *Drosophila* and mammals have shown that newly emerged male biased genes are preferentially found on the X chromosome [5,6]. In order to explain such discrepancy for male biased genes of different ages, we analyzed their expression along the three phases of male gametogenesis [3]. Differently than old genes, X-linked new genes are not depleted in the meiotic phase of spermatogenesis. In addition, we found an excess of autosomal new genes expressed in the post-meiotic phase. Therefore, we propose a model which combines the Faster-X hypothesis [7] and MSCI to explain differences on the distribution of male biased genes along evolution. At first, newly emerged genes are favored by male hemizygous expression in the X chromosome. As genes age, the MSCI machinery starts to act on their sequences until the X chromosome is no longer a favorable location for male biased genes.

[1]Betrán *et al.* 2002 Genome Res., 12:1854-9.

[2]Sturgill *et al.* 2007 Nature, 450:238-41.

[3]Vibranovski et al. 2009 PLoS Genet., 5:e1000731.

[4]Wu and Xu. 2003 Trends Genet. 19:243-7

[5]Zhang *et al.* 2010 Genome Res., 20:1526-33.

[6]Zhang *et al.* 2010 PLoS Biol., 8: e1000494.

[7]Vicoso, Charlesworth 2006 Nat. Rev. Genet.,7:645–53.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES13

Evolution of the extracellular matrix for multicellularity

Kasey Swilley [1,*], Katherine Johnson [1], Bradley Olson [1]

[1]Biology, Kansas State University, Manhattan, United States

Abstract: The evolution of multicellularity is a major transition in the morphological organization of organisms, however, the molecular mechanisms important for this transition in any taxa are currently not well understood. In most taxa, the molecular signature of the transition to multicellularity is obscured by nearly a billion years of divergence. Multicellularity evolved recently in the volvocine algae, thereby preserving the molecular signature of this transition. The volvocine algae include members that span the range or morphological complexity from unicellular (e.g. *Chlamydomonas*) to undifferentiated multicellular (e.g. *Gonium*), to species with differentiated tissues (e.g. *Volvox*). Importantly, the genomes of *Chlamydomonas* and *Volvox* have shown to be remarkably similar, suggesting the transition to multicellularity only requires the evolution of a few genes. To find genes important for multicellularity in undifferentiated multicellular G*onium pectorale,* we performed a genetic screen for unicellular mutants. From this we identified a mutant, *uc-1C7,* that is 99.6% unicellular. Resequencing its genome revealed that the causative mutation is in an ortholog of *GDT1,* which is conserved across all eukaryotes. GDT1 is localized to the trans-Golgi, where it plays a role in the proper glycosylation of proteins destined for the extracellular matrix. We found that the *uc-1C7* mutant is sensitive to detergent lysis consistent with defects in extracellular matrix assembly. When the *GDT1* ortholog from *Gonium* is expressed in unicellular *Chlamydomonas,* it causes a multicellular gain-of-function phenotype. These results suggest that GDT1 is important for multicellularity by regulating the maturation of proteins destined for the extracellular matrix to promote cell-cell adhesion.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

OM-ES12

Humanized yeast as a platform to study gene family expansion and understand genetic variation

Riddhiman Garge [1,*], Jon Laurent [1], Aashiq Kachroo [1], Edward Marcotte [1]

[1]Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, United States

Abstract: All presently existing species on the planet arose from a common ancestor suggesting that a substantial fraction of genes across species have common origin. *Saccharomyces cerevisiae* and *Homo sapiens* are separated by a billion years of evolution, yet a recent study from our lab demonstrated the replaceability of nearly half(~47%) of the tested essential yeast genes by their corresponding human ortholog(s) showing deep functional conservation of genetic networks in highly diverged organisms. Surprisingly, sequence similarity did not predict cross-species compatibility. However, replaceablility tended to be modular, i.e genes belonging to common pathways or complexes were similarly replaceable or not. To further understand replaceability of expanded gene families we are focussing on tubulins, the constituents of microtubules, that have expanded in the human, but contain one or two orthologs(in the case of the α-subunit) in the yeast lineage. Despite its functional conservation among eukaryotes, the rationale for its expansion is unknown. A wide range of diseases associated with different tubulin isoforms suggest varying functions across these isotypes. Humanized yeast thus supply a platform to study individual human genes in expanded families and learn how their genetic variation leads to disease. I will discuss the results from our tubulin humanization experiments including using humanized yeast as tools for high-throughput allelic variant and drug discovery screens.
Keywords- Humanization, high-throughput genetic variation screening, genetic networks

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Evolutionary systems biology of cells

POA-34

Hierarchical tissue organization as a general mechanism to limit the accumulation of somatic mutations

Gergely Szollosi [1,*], Imre Derenyi [2]

[1]MTA-ELTE "Lendulet" Evolutionary Genomics Research Group, [2]Dept. of Biological Physics, Eötvös University, Budapest, Hungary

Abstract:

How can tissues generate large numbers of cells, yet keep the divisional load (the number of divisions cell lineages) low in order to curtail the accumulation of somatic mutations and reduce the risk of cancer? To answer the question we consider a general model of hierarchically organized self-renewing tissues and show that the lifetime divisional load of such a tissue is independent of the details of the cell diffe entiation processes, and depends only on two structural and two dynamical parameters.
Our results demonstrate that a strict analytical relationship exists between two seemingly disparate characteristics of self-renewing tissues: divisional load and tissue organization. Most remarkably, we find that a sufficient number of progressively slower dividing cell types can be almost as efficient in minimizing the divisional load, as non-renewing tissues. We argue that one of the main functions of tissue-specific stem cells and differentiation hierarchies is the prevention of cancer.

Derenyi, I & Szollosi, GJ Hierarchical tissue organization as a general mechanism to limit the accumulation of somatic mutations. Nature Communications, in press (2017)https://doi.org/10.1101/098129

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Extracting insights from personal genomic data

OM-EI3

Estimation of nucleotide- and allele-specific selection coefficients for personal genomics using deep learning and population genetics

Yi-Fei Huang [1,*], Adam Siepel [1]

[1]Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, United States

Abstract: A central question in human genomics is understanding the functional, clinical, and evolutionary significance of variants identified by genome sequencing studies. Recently, several computational methods have been developed to estimate the strength of negative selection on genomic sequences by integrating numerous weak predictors of evolutionary constraints, such as conservation scores, histone modifications, and chromatin accessibility. Methods of this kind, such as LINSIGHT, fitCons, and CADD, not only provide insights into evolutionary constraints but they also help to prioritize putative disease variants for follow-up study. However, none of the existing methods is able to estimate selection coefficients, the most interpretable measures of natural selection. On the other hand, statistical methods based on the Poisson random field model in population genetics have been widely used to estimate distributions of selection coefficients for pre-identified genomic regions, but these methods are unable to estimate the effects of individual mutations. **Here, we describe a novel statistical framework, DeepINSIGHT, that unifies methods for variant prediction and estimation of distributions of selection coefficients, providing estimates of selection coefficients for all possible point mutations in the human genome.**

In particular, we formulate the estimation of selection coefficients as a regression problem in which the covariates are genomic features and the response is the observed derived allele frequency. DeepINSIGHT employs a deep-learning strategy to solve this regression problem. In DeepINSIGHT, the log likelihood function of the Poisson random field model is used in place of conventional cost functions used in classic deep learning models. This evolution-based cost function allows DeepINSIGHT to infer selection coefficients in a nucleotide- and allele-specific manner. In addition, as a neural network model, DeepINSIGHT is able to learn the potentially complicated nonlinear relationship between selection coefficients and genomic features. We apply DeepINSIGHT to a large number of genomic features and the high-coverage 1000 Genomes data and show that it produces highly accurate estimates of variant-specific selection coefficients, unmatched by any existing computational methods. Using known disease variants from public databases, we show that DeepINSIGHT is a powerful method both for obtaining insights into natural selection and for the prioritization of functional variants.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Extracting insights from personal genomic data

OM-E12

Human demographic history impacts genetic risk prediction across diverse populations

Alicia Martin [1,*], Christopher Gignoux [2], Raymond Walters [1], Genevieve Wojcik [2], Duncan Palmer [1], Benjamin Neale [1], Simon Gravel [3], Mark Daly [1], Carlos Bustamante [2], Eimear Kenny [4]

[1]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, [2]Genetics Department, Stanford University, Stanford, United States, [3]Department of Human Genetics, McGill University, Montreal, Canada, [4]Department of Genetics and Genomic Sciences, Mt. Sinai School of Medicine, New York, United States

Abstract: The vast majority of genome-wide association studies are performed in individuals of European descent, and their applicability to other populations is dependent on many factors (e.g. linkage disequilibrium, allele frequencies, and genetic architecture). As medical genomics studies become increasingly large and diverse, demographic models provide a critical lens into complex trait studies, for example informing the transferability of disease risk captured by European GWAS in understudied populations. We examined this transferability using published summary statistics for several well-studied traits and diseases and identified directional inconsistencies in all scores. These inconsistencies indicate that less of the heritable variation is explained in non-European cohorts. To gain deeper quantitative insights into GWAS transferability, we developed a complex trait coalescent-based simulation framework recapitulating demographic parameters in an Out-of-Africa model. We consider the effects of polygenicity, causal allele frequency divergence, and heritability. As expected, correlations between true and inferred risk are typically highest in the population from which summary statistics were derived. We demonstrate that scores inferred from European GWAS are biased in other populations even when choosing the same causal variants, and that biases in any direction are possible and unpredictable. We have also implemented novel methods that incorporate linkage disequilibrium into the correction of effect size estimates to improve cross-population transferability. Our work cautions that summarizing findings from large-scale GWAS may have limited portability to other populations using standard approaches, and highlights the need for the adoption of improved polygenic risk methods and the inclusion of more diverse individuals in medical genomics.

Expanded summary*: To date, GWAS have been performed opportunistically in primarily single-ancestry European cohorts, and an open question remains about their biomedical relevance for disease associations in other ancestries. GWAS have yielded tens of thousands of common genetic variants significantly associated with human medical and evolutionary phenotypes, most of which have replicated in other ethnic groups. However, GWAS are optimally powered to discover common variant associations, and the European bias in GWAS results in associated SNPs with higher minor allele frequencies on average compared to other populations. The predictive power of GWAS findings and genetic diagnostic accuracy in non-Europeans are therefore limited by population differences in allele frequencies and linkage disequilibrium structure. For example, a previous study showed that the accuracy of breeding values and genomic prediction decays approximately linearly with increasing divergence between the discovery and target population.

The dawn of the GWAS era saw limited success in identifying genome-wide significant loci associated with disease, and a major endeavor to better understand the genetic architecture of complex traits emerged. The peaks that met genome-wide significance typically did not explain a significant fraction of the phenotypic variance, and a major goal to estimate how many more signals remained yet to be discovered arose; this objective ushered in a wave of methodological development in heritability, linear mixed models, and polygenic risk prediction. Numerous complex traits have been studied with cohort sizes in the hundreds of thousands, and yet in each case there are many more signals that significantly improve prediction accuracy than meet genome-wide significance. Recently, several methodological advancements to the standard polygenic risk approach have been undertaken that increase the heritable variation explained. We have implemented a Kalman filter analogous to a continuous state Hidden Markov Model that incorporates linkage disequilibrium information to adjust effect size estimates, such that estimates more closely mirror causal (and therefore transferable) effect sizes.

We have shown that GWAS biases and demographic history have a striking impact on genetic risk prediction; for example, a previous study calculated polygenic risk scores for schizophrenia in East Asians and Africans based on GWAS summary statistics derived from a European cohort, and found that prediction accuracy was reduced by more than 50% in non-European populations. To disentangle the role of demography on polygenic risk prediction derived from single-ancestry GWAS, we designed a novel coalescent-based simulation framework reflecting modern human population history and show that polygenic risk scores derived from European GWAS are biased when applied to diverged populations. Specifically, we identify reduced variance in risk prediction with increasing divergence from Europe reflecting decreased overall variance explained. Our results highlight the need for the inclusion of more diverse populations in GWAS as well as wider adoption of genetic risk prediction methods that improve transferability across populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Extracting insights from personal genomic data

OM-EI6

Testing the generalizability of current GWAS findings within the multi-ethnic PAGE II Study

Genevieve Wojcik [1,*], Misa Graff [2], Jeff Haessler [3], Gillian Belbin [4], Carlos Bustamante [1 5], Christopher Gignoux [1], Eimear Kenny [6] and PAGE-II Study

[1]Genetics, Stanford University, Stanford, [2]Epidemiology, University of North Carolina, Chapel Hill, [3]Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, [4]Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, [5]Biomedical Data Sciences, Stanford University, Stanford, [6]Genetics and Genome Science, Icahn School of Medicine at Mount Sinai, New York, United States

Abstract: Over the past decade, genome-wide association studies (GWAS) have had enormous success in identifying disease loci. However, the majority of this research has prioritized populations of European ancestry. Key questions remain regarding generalizability in other populations, pointing ultimately to what we have yet to learn from trans-ethnic studies. Here we address questions of generalizability and population specificity across diverse populations at each step of the GWAS pipeline; array design, imputation, and association. We leverage data from >50,000 individuals linked to >100 traits in the PAGE-II Study of Hispanic/Latino, Asian, Hawaiian, and African-American ancestry, as well as a global reference panel including HGDP and indigenous African and American groups. We describe a novel multi-population tag SNP selection pipeline, which we applied to develop Illumina's Infinium Multi-Ethnic Genotyping Array (MEGA). This tag SNP selection scheme improved imputation accuracy globally, with meaningful power gains in association studies, particularly for rare variants. We infer an extraordinary breadth of population structure, admixture, and differential relatedness, allowing us to genetically characterize historically admixed populations and identify novel ancestry-specific associations. We also conducted admixture-based analyses for a set of common phenotypes, highlighting the utility of genetic ancestry for phenotype-wide association studies (PheWAS). Finally, we explore genetic relationships between and within traits and populations, including characterizing patterns of replication in previously reported GWAS signals. In summary, this work highlights the importance of genomics within diverse populations to elucidate potential disease loci in a meaningful manner, and lays the groundwork for the next generation of large-scale medical genetics studies.

Expanded summary*: The success of the first decade of genome-wide association studies (GWAS) has been largely predicated on a model of discrete, homogeneous populations—typically of European descent. However, within the United States the majority of disease burden disproportionately affects minority populations. Understanding the genetic architecture of complex disease in diverse ethnic populations is essential to decoding disease pathogenesis and the reduction of health disparities. As we scale to large-scale genomics in diverse populations, we need to address the reality where most populations are on a heterogeneous continuum and can represent a mixture of ancestries. Modeling this process requires improved tools and methods for discovery. Additionally, there are questions of the generalizability of these results to other populations which must be explored.

The Population Architecture using Genomics and Epidemiology (PAGE)-II Study was formed as part of the effort to address this knowledge gap. Drawing from multiple large-scale cohorts and a metropolitan biobank, a total of 50,000 samples were included from non-European descent groups. Individuals self-identified as Hispanic/Latino, African-American, Asian, Native Hawaiian, or Native American. A major challenge with the multi-ethnic study design was the lack of genotyping array that would adequately capture variation across all populations simultaneously. As a collaboration with both academic and industry partners, we designed Illumina's Infinium Multi-Ethnic Genotyping Array (MEGA). A novel pipeline was developed to select tag SNPs for the GWAS scaffold for optimal imputation accuracy across all ethnicities. The now

commercially-available array represents within the field a step forward for large-scale genetic studies which include transethnic study design.

The PAGE-II study populations were primarily admixed, self-identified as Hispanic/Latino (~22,000) and African-American (18,000). To characterize within- and between-group differences, we estimated global ancestry with principal components analysis (PCA) and ADMIXTURE. Within Hispanic/Latino individuals, we see heterogeneity in the proportion of African, European, and Native American ancestry. These differences highlight the importance of population granularity, despite many studies assuming homogeneity. We are also able to replicate previous findings showing a large founder ancestry component in Caribbean populations, most notably in Puerto Rico.

Once we were able to estimate the admixture present in these populations, it was important to leverage their unique genetic architecture for complex trait mapping. PAGE-II has harmonized over 200 phenotypes across the study populations, providing an opportunity to explore the relationship of genetic variation and ancestry with disease. By evaluating the association between continental ancestral components and numerous phenotypes, we are able to elucidate the underlying role of genetic ancestry in the heterogeneity of risk between ethnic groups. We also examined the influence of genetic ancestry on the transferability of GWAS discoveries to external populations by quantifying replication rates of known GWAS signals within PAGE-II. Additionally, we explore genetic correlations between traits and the possible relationship with ancestry. This work is essential for understanding the transferability of results from the predominantly European-descent GWAS catalog to the populations where the majority of chronic disease burden lies.

In summary, by characterizing the genetic ancestry of diverse populations and leveraging their unique genetic architecture for trait mapping, we are able to bring light to the complex ways ancestry affects human health. This work is a crucial step towards understanding and eliminating health disparities.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Extracting insights from personal genomic data

POA-166

Exploring the Effects of Ancestral Genetic Composition on African American Identity

Fatimah Jackson [1,*], Christopher Cross [2], Latifa Jackson [3]

[1]Biology, W. Montague Cobb Research Laboratory, [2]Anatomy, [3]Pediatrics, National Human Genome Center, Howard University, Washington, United States

Abstract: Genetic analyses have been invaluable in elaborating and often clarifying population history. Studies of the genetics of African diasporic populations are in their infancy, largely because of the lack of interest by majority population researchers, limited historical knowledge of the origins and dispersions of New World Africans, and a paucity of meaningfully collaborative teams of researchers exploring population substructuring in the African Diasporas. In our study, self-identified African Americans Biology majors (N=40 individuals, Range= 18-23 years) at Howard University were provided with a pre-survey questionnaire that asked about who they thought they were descended from. Saliva samples were then submitted to 23andMe for ancestral genomic evaluations. Participants then completed a post-survey questionnaire after receiving their results. Pooled results of the genomic evaluations were discussed and the historical context for findings presented. Participants fell into three main categories of response: those whose ancestral genetic identity matched with expectations from their family stories and their phenotype; those whose ancestral genetic results did not confirm family oral histories but were reconciled through reinterpretation of the results; and those participants whose ancestral genetic results provided an unexpected/unwelcome insight into their past. In this latter case, genetic information was rejected. Overall the results of the ancestral genetic tests increased participant interest in the “why and where” of particular genetic variants and information on coalescence times for particular haplotypes was internalized by many of the participants. Counterintuitively, participants were more receptive to the genetic ancestry results when they knew less information about their immediate biological lineage.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Extracting insights from personal genomic data

POB-408

INSIGHT INTO CETACEAN HYPOXIA ADAPTATION FROM MOLECULAR EVOLUTION OF ENERGY METABOLISM RELATED GENES

Ran Tian [1,*], Shixia Xu [1], Guang Yang [1]

[1]Nanjing Normal University, nanjing, China

Poster: Abstract

Oxygen is indispensable for energetic metabolism, and insufficient oxygen supply will result in deficit of energy demands of cells. It has been suggested that physiological modifications of energy metabolism allowed cetaceans to conquer the hypoxic-diving niche. However, the molecular mechanisms underlying the energy metabolism remain unknown. Thus, we examined 194 nuclear and mitochondrial genes involved in four energetic metabolism pathways in an attempt to gain insights into the evolution of energy metabolism underlying hypoxia adaptation. The most positively selected genes (PSG) were enriched in citrate cycle (TCA cycle) for cetaceans, suggesting an enhanced capability for aerobic metabolism of cetaceans during most natural dives. Moreover, six rate-controlling enzyme genes (LDHA, LDHD, PC, PCK1, FBP1, and GPI) involved in gluconeogenesis shown cetacean-specific amino acid changes, and higher lactate dehydrogenase (LDHA) enzyme activity were detected in cetaceans, both suggesting that cetaceans have enhanced the reconversion ability of lactate to glucose with long diving duration. Besides, evidence for positive selection was also detected in other hypoxia tolerant lineages; and but divergences of positively selected genes (PSGs) number, distribution of PSGs along each pathway, and species-specific PSGs indicated genetic discrepancy between hypoxia tolerant species adapted to unique habitats. Additionally, convergent/parallel amino acid substitutions were also observed between hypoxia tolerant mammals, several of which were in genes evolving under positive selection, and genes that perform functions that are essential for hypoxia tolerance. This results suggested that convergent molecular evolution might underline hypoxia tolerant mammals to adapt to hypoxic stress, although these mammals highly diverged with specific ecological /physiological adaptation.

Key words: Hypoxia tolerance, energy metabolism, adaptive evolution, positive selection, convergent evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Extracting insights from personal genomic data

OM-EI4

Clinical Characterization of over 30,000 clinically relevant variants in 133 global populations

Elena Sorokin [1,*], Gillian Belbin [2], Genevieve Wojcik [1], Carlos Bustamante [1], Chris Gignoux [1], Eimear Kenny [2] and The Population Architecture using Genomics and Epidemiology (PAGE)-II Study

[1]Genetics, Stanford University, Stanford, CA, [2]Genetics and Genome Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States

Abstract: The profusion and population-specificity of rare variants in human populations make questions of causality, penetrance and expressivity of clinically relevant variants (CRVs) difficult to ascertain, even within well-characterized disease genes. Therefore, characterization of CRVs in multi-ethnic cohorts is a critical step in advancing global precision medicine. We genotyped >30K CRVs on the Multi-Ethnic Genotyping Array in >52K participants representing 133 global populations as part of the PAGE-II study. These include: 7,615 pathogenic sites, 2,553 sites within the 56 genes prioritized by the ACMG for reporting incidental findings, and 2,611 pharmaco-variants. We observed an average of 23 pathogenic variants per individual, higher within African descent populations and admixed American populations. We identify 4,513 ClinVar variants with MAF >1% in any population, including 157 pathogenic/pharmacogenetically-actionable. We observe high differentiation with 323 variants (136 in pharmacogenes) with Fst >0.4 between non-Europeans and Europeans. We genotyped the Mount Sinai Bio*Me* biobank allowing us to genetically identify and clinically characterize segregating CRVs in a large urban health system. We will highlight examples of some pathogenic CRVs, thought to be rare, which are segregating appreciably in NYC founder populations, and are linked to relevant clinical outcomes. We also elucidate the prevalences of well-characterized genetic disorders in diverse populations. For example, familial hypercholesterolemia (FH) has a surprisingly high carrier frequency of 1:49 within African-Americans, almost 5-fold more common compared to European Americans, and driven by a single variant in PCSK9. Patterns of drift in human populations refine and empower our understanding of CRVs in multi-ethnic cohorts.

Expanded summary*: Despite major advances in uncovering the genetic basis of disease, genetic testing has not been broadly implemented in clinical practice. Among the hurdles to clinical implementation is the potential for misdiagnosis, since variants in medical databases were originally ascertained in European populations; and the need to identify at-risk individuals based on ancestry, so that appropriate tests can be routinely administered for preventive screening. Both challenges can be mitigated by diversity: genotype-phenotype association studies in large, diverse cohorts can tease apart ancestry-specific variation from truly deleterious variation, and studies in diverse cohorts can establish statistical priors in order to identify individuals who who have increased chance for a given disorder or adverse drug response, based on ancestry. In my work with the Kenny and Bustamante groups at Mount Sinai and Stanford respectively, we are leveraging population genetics in diverse populations to address both of these hurdles regarding the uptake of genomic medicine.

One major problem with medical variant databases such as ClinVar and its antecedent, OMIM, is that many variants were annotated based on clinical sequencing in small studies and European cohorts. A recent report described reclassified variants in two genes that cause hypertrophic cardiomyopathy, a life-threatening and often fatal heart condition; the variants of interest were exceedingly rare in European-descent populations, yet relatively common in populations of African ancestry (Manrai et al *NEJM* 2016). Colleagues in the Kenny and Bustamante labs developed a genotyping platform, the MEGA array, aimed at multi-ethnic mapping within the Population Architecture for Genomics and Epidemiology (PAGE)-II study, which subsequently genotyped over 50,000 individuals including African-Americans, Hispanics, East Asians, Native Hawaiians, Native Americans and South Asians. The MEGA array was designed to include over 34,000 clinically relevant variants from OMIM, ClinVar, PharmGKB and other clinical databases. Analyzing this data, I identified over 200 variants annotated as pathogenic, conflicted, or uncertain significance, with a risk allele frequency over 1% in one or more large (n>500) PAGE populations. We have also identified 136 pharmacogenetic variants with substantially differentiated minor allele frequency (Fixation index >0.4) between PAGE populations and a cohort of Europeans. We are bridging with the ClinGen consortium to disseminate PAGE allele frequency data to medical geneticists, clinicians and the public and are currently working with ClinGen software engineers to design a web interface for data sharing.

For the second challenge, identification of at-risk populations, I have been closely collaborating with the Eimear Kenny laboratory, which is embedded within the Mount Sinai health system and has access to electronic health records from over 12,000 Mount Sinai participants who were genotyped on MEGA and within PAGE. The Kenny lab has developed a Phenome-wide association (PheWAS) platform using ICD9 medical billing codes. As an example, we have identified a risk variant for familial hypercholesterolemia (FH) that is present in African-Americans at 0.98%, and is signfiicantly associated with higher blood LDL levels, the signature of FH. This is an example of a genetic disorder that may be up to five times more prevalent in a non-European population compared to Europeans, where it has been traditionally studied. Using this linked genotype-phenotype data within the diverse Mount Sinai biobank, we can start to identify populations with increased prevalence for FH and other common genetic disorders. Ultimately, these and other efforts will reduce healthcare costs through prediction of adverse drug response and early detection of disease.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Extracting insights from personal genomic data

OM-EI1

Past and future of prediction of medical phenotypes from genome sequence

Andrew G. Clark [1,*]

[1]Molecular Biology and Genetics, CORNELL UNIVERSITY, Ithaca, United States

Abstract: The ability to predict agriculturally important phenotypes from genome-wide SNP genotypes in domesticated plants and animals is revolutionizing agriculture, and if genomic prediction worked as well in humans, it would revolutionize medicine. The current state of the art in human phenotype prediction is to apply agriculture-based approaches such as PRS (polygenic risk score) and G-BLUP, as derived from general linear mixed models. Successful prediction of phenotype in agriculture generally entails prediction of the mean of a large number of offspring (i.e. the breeding value), and this is inherently easier to predict than phenotypes of single individuals. When the number of predictors (SNPs) exceeds the sample size, power to infer interactions (epistasis and GxE) gets killed by the number of tests, and even if there is extensive epistasis, the models rarely can detect it. The vagaries of individual environmental experience average out when predicting the mean of a progeny set, but such individual exposures do matter for individual prediction. We have seen spectacular progress recently in the application of a form of machine learning called deep learning, with tools that perform surprisingly well at tasks like image recognition. It seems likely that tools like this will be able to beat all other comers, especially if they can successfully integrate information from Electronic Medical Records, including family history and environmental exposures (drugs, physical activity, etc.). These approaches will be explored with twin and GWAS data, and implications of model-free predictions surpassing mechanism-based understanding will be discussed.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Extracting insights from personal genomic data

POA-165

Tracking the origins and migration pattern of metastatic cancer

Bingjie Chen [1,*], Yongsen Ruan [1], Qingjian Chen [1], Zuyu Yang [2], Xu Shen [1], Chung-I Wu [1 2 3]

[1]College of Ecology and Evolution, Sun Yat-Sen University, Guangzhou, [2]Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China, [3]Department of Ecology and Evolution, University of Chicago, Chicago, United States

Abstract: Reconstructing the evolutionary process of tumor growth and migration is important for understanding cancer basic biological problems, which may also provide new insight into therapeutic instructions. At present, the underling principles involving temporal, spatial and quantitative questions of metastatic seeding at distant sites from primary tumor are remain controversial. In this study, we use primary(P) and multiple metastases(Ms) in distant organ from a chemotherapy-naive cancer patient. Through high-density three-dimensional clonal sampling, whole genome(78X) and deep target sequencing(600X), we obtain mutation information including single-nucleotide variations (SNVs) and structure variation(SV) of nearly 300 sampling sites of P and Ms. After anatomic location reconstruction and phylogeny analysis, we detect parallel distinct waves disseminate from two different clones to distant metastases and we also find the signal of cross-mutation between different metastases which may be the result of multiple founder clones or metastasis-to-metastasis spread. Besides, we are also trying to figure the difference of evolutionary driving force and clonal structure between primary and metastatic tumor.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Extracting insights from personal genomic data

POA-164

Effective migration maps using 770,000 individuals reveal geographical barriers and migration patterns

Shiya Song [1,*], Ariel Anderson [2], Nathan Berkowitz [1], Jake Byrnes [1], Ross Curtis [2], Eyal Elyashiv [1], Daniel Garrigan [1], Ahna Girshick [1], Julie Granka [1], Harendra Guturu [1], Natalie Myres [2], Keith Noto [1], Kristin Rand [1], Oren Schaedel [1], David Turissini [1], Yong Wang [1], Ben Wilson [1], Eurie Hong [1], Catherine Ball [1], Ken Chahine [2]

[1]Ancestry.com DNA, LLC, San Francisco, [2]Ancestry.com DNA, LLC, Lehi, United States

Abstract: Geographic distances often cause genetic similarities to decay. Heterogeneity in the rate of decay is caused by various factors affecting gene flow. Understanding gene flow barriers or corridors is crucial to identify fine scale population structure and can yield insights on evolutionary processes that shape present day genetic diversity. A newly published method EEMS (Estimated Effective Migration Surfaces) models the relationship between genetics and geography and produces a visual representation of population structure highlighting potential regions of historic gene flow (Petkova *Nature* 2016). Using genome-wide SNP data from over 770,000 AncestryDNA customers who consented to research, we applied EEMS and estimated effective migration rates among over 63 geographically localized populations, representing continental, sub-continental and fine-grained regional scales. We identified several known geographic barriers and regions showing high connectivity. We further investigated interesting migration patterns and correlated these patterns with known pedigree information and historical records.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Extracting insights from personal genomic data

POA-168

Multivariate multiscale impacts of genetic variants on gene expression variability in humans

James Cai [1,*]

[1]Texas A&M University, College Station, United States

Abstract: Increasing evidence shows that phenotypic variance is genetically controlled, but the underlying mechanisms of genetic control over the variance remain obscure. Here, we conducted variance-association mapping analyses in humans to identify expression variability QTLs (evQTLs), i.e. genomic loci associated with gene expression variance. We discovered abundant common genetic variants altering gene expression variance via two distinct modes of action---epistasis and destabilization. The impact of common variants on gene expression variability is detectable in population as well as single-cell levels. Furthermore, rare and low-frequency variants cause variable gene expression. To assess the impact of rare and low-frequency variants on gene expression variability, we established a novel analytical framework based on multivariate outlier detection and identified rare and low-frequency variants causing aberrant gene expression. Together, our findings contribute to the understanding of the mechanisms of genetic control over phenotypic variance and may have implications for the development of variance-centered analytic methods for quantitative trait mapping.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Extracting insights from personal genomic data

OM-EI5

Ascertainment bias creates the illusion of genetic health disparities and limits the generalizability of results across populations

Michelle Kim*, Kane Patel, Andrew Teng, Ali Berens, Joe Lachance

Abstract: Here, we integrate whole genome sequence data from the 1000 Genomes Project with the NHGRI-EBI GWAS Catalog to predict hereditary disease risks across the globe. We find substantial regional differences in risk allele frequencies across populations, including elevated frequencies in Africa (approximately 1.3% high risk frequencies). These patterns are due to multiple forms of ascertainment bias. First, most GWAS use samples of European ancestry, and this causes rare variants in other populations to be missed. Second, commonly used genotyping arrays contain an excess of intermediate frequency alleles. Genotyping arrays are also enriched for SNPs that have elevated frequencies of derived alleles in non-African populations. Third, results are less likely to be generalizable across populations if these populations do not share recent common ancestry. We find substantial bias in genetic risk score predictions for ancestral and derived alleles: risks are underestimated in study populations if ancestral alleles tag diseases and risks are overestimated in study populations if derived alleles tag diseases. By simulating GWAS that use different numbers of cases and controls, study populations, and genotyping technologies we are able to successfully capture general patterns of ascertainment bias found in real-world datasets. Disease associations do not always generalize well across populations and studies, and it is important to correct for biases when estimating hereditary disease risks in diverse human populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM4

Isolating parent-of-origin effects causing hybrid inviability in Mimulus

Elen Oneal [1,*], Miguel Flores-Vergara [2], Robert G. Franks [2], John Willis [1]

[1]Department of Biology, Duke University, Durham, [2]Plant and Microbial Biology, North Carolina State University, Raleigh, United States

Abstract: Hybrid seed lethality is a common outcome of hybridization between closely related plant species, suggesting that genetic mechanisms governing seed development may diverge rapidly. Interploidy crosses in *Arabidopsis thaliana* have revealed that dosage imbalances in genes critical to normal endosperm development often underlie hybrid seed failure, and moreover, that these genes frequently exhibit differential expression of alleles based on parental origin (genomic imprinting). Seed failure in crosses between diploid plants may similarly result from mis-expression of evolutionarily divergent alleles in the endosperm of hybrid seeds. We examine this possibility using the diploid species pair *Mimulus guttatus* and *M. nudatus*, which are recently diverged and yet nearly completely reproductively isolated. Hybrid seed exhibit impaired endosperm development and are almost always inviable. We utilize a unique triangulated crossing design that allows us to bypass the hybrid barrier and alternate the parental source of alleles, enabling the detection of QTL associated with hybrid inviability and parent-of-origin effects. Preliminary mapping results reveal two major QTL that differ in their effects when they are maternally inherited vs. paternally inherited. Moreover, each QTL contains promising candidate loci, including two homologs of the PRC2 complex, a major regulator of imprinted gene expression in endosperm in *Arabidopsis*. Together with transcriptome data from whole *M. guttatus* seed and endosperm, our quantitative mapping data provide valuable insights into the developmental and evolutionary mechanisms responsible for hybrid seed inviability in the *Mimulus* species complex.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

POB-193

Detecting introgressed loci using machine learning

Daniel Schrider*, Andrew Kern

Abstract: Speciation with gene flow appears to be common, and knowledge of which loci have and have not experienced such gene flow can elucidate the genetic basis of hybrid incompatibility. Moreover adaptive introgression, in which beneficial alleles cross species boundaries, appears to be commonplace in nature. Thus it is crucial to develop the statistical machinery required to uncover which genomic regions have recently acquired haplotypes via introgression from a sister species. We developed a novel machine learning framework, called FILET (Finding Introgressed Loci via Extra-Trees) capable of revealing genomic introgression with far greater power than competing methods. FILET works by combining information from a number of population genetic summary statistics, including several new statistics that we introduce, that capture patterns of variation across two populations. We show that FILET is able to determine at which loci recent gene flow has occured among related species with unparalleled accuracy, and in most situations can correctly infer which population was the donor and which was the recipient. FILET can also detect gene flow from an unsampled "ghost" population into a sampled "sink" population, with remarkable sensitivity and specificity. We therefore applied FILET to address the question of whether human populations in Africa have been the recipients of gene flow from an archaic hominin, and find numerous loci showing strong signatures of such introgression. These results provide confirmation of previous studies suggesting gene flow from an unknown an archaic hominin donor population into African populations within the Late Pleistocene.

Expanded summary*: Research Significance and Summary – Daniel Schrider

There is a growing body of evidence that, following speciation, the two incipient species may continue to sporadically interbreed even over a fairly long evolutionary timescale. Indeed several fascinating examples of such hybridization have occurred in recent human evolution, including evidence that roughly 2% of non-African humans' genetic material derived from Neanderthals [1], and that perhaps an even larger fraction of Oceanic individuals' DNA is of Denisovan ancestry [2]. Examining the genomic landscape of introgression across species can aid efforts to uncover the basis of genetic incompatibilities between species as well as local adaptation, as cross-species gene flow of such loci may be selected against. On the other hand, introgression from a related species can be an important source of beneficial alleles. For example, a variant of the *EPAS1* gene found in Tibetans is adaptive at high altitudes and was donated to modern humans from Denisovans [3]. Thus, in order to gain a better understanding of the processes of hybridization, speciation, and local adaptation, we require methods that can determine which regions of the genome have recently introgressed from one population to another.

When a haplotype migrates from one population to another, patterns of genetic variation at the introgressed locus are skewed in several notable ways. First, an increase in linkage disequilibrium (LD) is expected within the recipient population, along with an increase in genetic diversity in this population, an increased number of alleles shared across populations, and various other patterns produced by a genealogy that includes a recent migrant from one population to another. Here I describe FILET (Finding Introgressed Loci using Extra-Trees classification), a novel method that seeks to simultaneously capture all of these signatures in order to detect introgression with greater sensitivity and specificity. FILET leverages a machine learning approach called Extra-Trees classification, an extension of random forests [4] which discriminate among two or more classes of data on the basis of a large number of dimensions or "features." In the case of FILET, the classes correspond to "no introgression," "introgression from population 1 to 2," and "introgression from population 2 to 1." The features correspond to a large set of population genetic summary statistics (e.g. F_{ST}, features of the identity-by-state tract length distribution, measures of LD in each population, etc).

Using coalescent simulations we show that FILET vastly outperforms competing methods under a wide variety of scenarios (i.e. a wide range of timings of population splits and subsequent hybridization events). Moreover, we demonstrate that FILET can infer the direction of introgression event with remarkable accuracy. Lastly, we extend FILET to the problem of identifying loci donated from an unsampled "ghost" population, combining both previously devised and novel statistics as our features, and again achieve excellent accuracy on simulated data. Together, the success of these two flavors of FILET vastly broadens the range of scenarios in which loci derived from a diverged population can be identified. In light of the unparalleled success of this approach, I will argue that population

genetic inference is due for a paradigm shift away from methods relying on only a single test or statistic toward methods that incorporate multiple disparate summaries of the data (such as the machine learning approach discussed here).

After briefly describing FILET and characterizing its inferential power, I will share insights revealed by applying it to population samples from the recently diverged sister species *Drosophila simulans* and *Drosophila sechellia*, uncovering evidence of a substantial rate of gene flow from *simulans* to *sechellia*. I then apply the "ghost" population version of FILET to data from each African population sample from the 1000 Genomes project. It has been suggested that some sub-Saharan African populations contain DNA whose ancestry traces back to an unknown archaic hominin [5]. Our analyses reveals numerous loci that show strong signatures of recent introgression (within 100 kya) from a population that diverged from modern humans >500 kya. This result implies that the extent to which modern humans share ancestry with archaic hominin relatives is even greater than has been suggested by comparisons with ancient DNA samples.

References

1. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. Science. 2010;328(5979):710-22.
2. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010;468(7327):1053-60.
3. Huerta-Sánchez E, Jin X, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 2014;512(7513):194-7.
4. Breiman L. Random forests. Machine Learning. 2001;45(1):5-32.
5. Hammer MF, Woerner AE, Mendez FL, Watkins JC, Wall JD. Genetic evidence for archaic admixture in Africa. Proceedings of the National Academy of Sciences. 2011;108(37):15123-8.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

POB-190

Gene flow and selection between major clades of the Malawi cichlid adaptive radiation

Richard Durbin [1,*], Hannes Svardal [1], Milan Malinsky [2], Alexandra Tyers [3], Bosco Rusuwa [4], Eric Miska [5], George Turner [6]

[1]Wellcome Trust Sanger Institute, Cambridge, United Kingdom, [2]University of Basel, Basel, Switzerland, [3]University of Bangor, Bangor, United Kingdom, [4]Chancellor College, Zomba, Malawi, [5]Gurdon Institute, Cambridge University, Cambridge, [6]Bangor University, Bangor, United Kingdom

Abstract: Over the past few years we have sequenced over 300 fish from over 100 species of Lake Malawi haplochromine cichlids, providing extensive new insights into this most dramatic of recent vertebrate evolutionary radiations. All species are genetically close, with pairwise divergence typically between 0.1 and 0.25%, compared to heterozygosity between 0.05 and 0.15%. We see many signals of gene flow and locus-specific selection between genetically close but phenotypically divergent species. Here we focus on the Utaka, which form one of the five ecologically and genetically distinct subgroups of the Malawi radiation, comprising most but not all *Copadichromis* species. (As with some other genera, we find that *Copadichromis* is not monophyletic, with some species in the shallow benthic or sand-dweller subgroup.)

Using PCA, F statistics and related methods, we see a strong signal of gene flow from *Diplotaxodon*, pelagic plankton eaters mainly found in deep water, into Utaka, which are also plankton eaters, but found closer to the coast in shallower water. There is a weaker signal into some shallow benthic species, which on a PCA plot smear out towards Utaka and *Diplotaxodon*. We see signs of several large inversions (up to half a chromosome) between these species, within which there are distinct phylogenies from the genome-wide average, though the gene flow signals also exist outside these inversions. Surprisingly *Copadichromis trimaculatus* appears to lie genetically halfway between the Utaka *Copadichromis* and the shallow benthic *Copadichromis*, entirely inconsistent with the top level phylogeny.

Initial comparisons between *Copadichromis virginalis* and *Diplotaxodon limnothrissa* indicate that while there is average species separation Fst of 23.4%, at certain places in the genome there are spikes of high Fst over 90%, and similarly there appear to be intervals with very low Fst under 5%. More broadly we see higher differentiation between species in genes involved in retinal processing, the innate immune system, mitochondrial function, and a number of other pathways including some developmental pathways.

In January 2017 we collected approaching 1000 additional samples, focussing on Utaka and similar species both from a number of locations in southern Lake Malawi, and from the nearby much smaller, younger and shallower Lake Malombe, which is just downstream of the outflow of Lake Malawi. Lake Malombe has been severely overfished in the last 10-20 years and several species appear to mature at substantially smaller size there than in Lake Malawi. We hope to report on the analysis of these data at the meeting, focusing on signals of selection between the lakes and between species within Malawi.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM9

A wolf in sheep's clothing: how a killer selfish element was disguised as a developmental gene in C. elegans.

Alejandro Burga [1,*], Eyal Ben-David [1], Leonid Kruglyak

[1]UCLA, Los Angeles, United States

Abstract: Genetic incompatibilities restrict gene flow between populations, and can thus contribute to speciation. While genetic incompatibilities are widespread across the tree of life, in only a few cases the underlying genetic mechanisms have been dissected. We discovered a novel genetic incompatibility in the nematode *C. elegans*, which stems from the interaction between a maternal-effect toxin and a zygotically expressed antidote. We mapped the incompatibility to two linked genes that constitute a selfish element: *sup-35* and *pha-1*. *pha-1* was originally proposed to be a developmental gene required for differentiation of the pharynx, and *sup-35* was identified as a suppressor of *pha-1*. Our results show that the pharyngeal defects identified in embryos lacking *pha-1* are instead a direct consequence of *sup-35* toxicity. We characterized the global pattern of variation in the activity of the *sup-35/pha-1* selfish element and found that loss of its driver activity evolved independently at least twice. In addition, we found a large inversion in all strains carrying the active form of the element that could have contributed to its fixation by preventing recombination from decoupling the loci. Our results illustrate how the study of natural genetic variation can illuminate our understanding of gene function and suggest that other developmental genes identified using forward genetics may be selfish elements.

Expanded summary*: Broad Significance:

1. This is the first genetic and molecular characterization of a maternal-effect toxin causing hybrid inviability. Understanding how selfish element spread iin natural populations is not only important to understand the dynamics of speciation, but also crucial to better design synthethic driver elements for pathogen control.

2. Our study also illustrates how selfish elements can be easily mistaken for developmental genes and provides a cautionary tale for developmental biologist and geneticists.

Expanded Summary

Selfish genetic elements promote their own transmission while being neutral or detrimental to the fitness of the organism. In extreme cases, selfish elements promote their transmission by killing individuals that do not inherit them, leading to genetic incompatibilities between carriers and non-carriers (1). Our laboratory previously identified the only known genetic incompatibility in the nematode *Caenorhabditis elegans*. The incompatibility is caused by a selfish element composed of two tightly linked genes: *peel-1*, a sperm-delivered toxin, and *zeel-1*, a zygotically expressed antidote (2). In crosses between isolates that carry the element and ones that do not, the element acts in heterozygous males or hermaphrodites via paternal effect to kill the progeny that do not inherit it. Selfish elements are predicted to spread in natural populations, and consequently, there is significant interest in using synthetic forms of such elements to drive population replacement of pathogen vectors in the wild (3). However, despite the prominent role genetic incompatibilities play in genome evolution and their promise in pathogen control, the genetic mechanisms underlying most incompatibilities remain largely unknown.

We discovered a genetic incompatibility in *C. elegans* that is caused by an interaction between a maternally deposited toxin and a zygotically expressed antidote. A comparable element, *Medea*, has been previously described in the beetle *Tribolium*; however, the underlying genes in *Tribolium* remain unknown (4). In this study, we identified the genes underlying a *Medea-like* element: the maternal toxin *sup-35* and the zygotic antidote *pha-1*. The antidote, *pha-1,* was originally thought to be a developmental gene, in large part due to the specific pharyngeal defects observed in mutants (5). However, the precise role of *pha-1* in embryonic development remained elusive and controversial over the years. Our results indicate that *pha-1* pharyngeal defects are a direct consequence of *sup-35* toxicity. Consistent with this model, all *pha-1* associated defects can be rescued by mutations in *sup-35* and *sup-35 mutants* do not show any obvious deficiency compared to wild type worms in the laboratory. Furthermore, we have identified *sup-35* mutants among naturally occurring wild isolates. Taken together, these results indicate that *sup-35* and *pha-1* likely act as a selfish element spreading in *C. elegans* wild populations, and are not integral components of *C. elegans* embryonic development as originally suggested.

One important insight emanating from previous work in light of our results, is that the *sup-35/pha-1* element exerts its toxicity by recruiting genes that are directly involved in *C. elegans* development (6). The other two known suppressors of *pha-1* lethality, *sup-36* and *sup-37,* are essential for *sup-35* toxicity and are conserved in other nematodes. Interestingly, *sup-37* is required for normal pharyngeal pumping and promotes ovulation in the somatic gonad independently of *pha-1* function. Null *sup-37* mutants are inviable and undergo early larval arrest. However, a single missense and viable mutation in *sup-37* is sufficient to abolish *sup-35* toxicity. Together with the finding that SUP-37 physically interacts with SUP-35, this suggests that the *sup-35/pha-1* selfish element is hijacking a developmental pathway to kill those embryos that did not inherit the element. These observations can also explain the apparent paradox of a parentally deposited toxin that is suppressed by a zygotic antidote that is only expressed later. It also suggests that the specificity in the activity and expression of *sup-36* and *sup-37* may explain the pharyngeal phenotypes of *pha-1* mutants.

Our work highlights the importance of studying natural genetic variation for understanding gene function. Despite the indisputable value of a common reference strain, it has proved extremely difficult in the context of the N2 background alone to either confirm or rule out *pha-1* as an essential component of *C. elegans* embryonic development. The study of other wild isolates has made possible our characterization of *sup-35/pha-1* as a selfish element. Our results also suggest that selfish elements conferring genetic incompatibilities may be more common than previously thought, and that some of them may be hiding in plain sight.

References

(1) G. D. Hurst, J. H. Werren, The role of selfish genetic elements in eukaryotic evolution. *Nat. Rev. Genet.* 2, 597–606 (2001).

(2) H. S. Seidel, M. V Rockman, L. Kruglyak, Widespread genetic incompatibility in C. elegans maintained by balancing selection. *Science*. 319, 589–94 (2008).

(3) C.-H. Chen *et al.*, A Synthetic Maternal-Effect Selfish Genetic Element Drives Population Replacement in Drosophila. *Science*. 316 (2007).

(4) R. Beeman, K. Friesen, R. Denell, Maternal-effect selfish genes in flour beetles. *Science*. 256 (1992).

(5) H. Schnabel, R. Schnabel, An Organ-Specific Differentiation Gene, pha-1, from Caenorhabditis elegans. *Science*. 250 (1990).

(6) D. S. Fay *et al.*, A Regulatory Module Controlling Pharyngeal Development and Function in Caenorhabditis elegans. *Genetics*. 191 (2012).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM1

Mechanisms and evolution of imprinting in plants

Mary Gehring*

Abstract: In plants, imprinting occurs in the endosperm, an essential seed tissue formed from one of two fertilization events that characterize flowering plant reproduction. The endosperm is triploid, with two maternally and one paternally inherited genomes. Most genes are expressed in a maternal: paternal ratio of 2:1, but 100+ genes are imprinted, meaning they are expressed preferentially either from the maternally or the paternally inherited alleles. This talk will focus on our latest understanding of the role of DNA methylation dynamics, transposable elements, and various epigenetic pathways in establishing and maintaining imprinted expression, and will discuss insights derived from comparative evolutionary studies of imprinting.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM8

Dissecting the special role of sex chromosomes in Drosophila speciation: genetic mapping of hybrid male sterility genes

Colin Meiklejohn*, Emily Landeen [1], Kathleen Gordon [2], Shelby Biel, David Stern [3], Daven Presgraves [4]

[1]Department of Integrative Biology, University of California Berkeley, Berkeley, [2]School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, [3]HHMI, Janelia Research Campus, Ashburn, [4]Department of Biology, University of Rochester, Rochester, United States

Abstract: In most animals with heteromorphic sex chromosomes, the evolution of X-linked hybrid sterility is an early step in the evolution of complete reproductive isolation, as genetic factors causing hybrid sterility in the heterogametic (XY) sex accumulate faster on the X chromosome than the rest of the genome. Why the X chromosome has unique roles in the rapid evolution of gametogenesis and speciation is unclear. To determine the genetic, developmental, molecular, and evolutionary reasons why the X chromosome is a hotspot for the accumulation of hybrid sterility, we have generated a high-resolution genetic map of X-linked hybrid male sterility factors between two sister species, *Drosophila mauritiana* and *D. simulans.* This map comprises 482 overlapping introgressed segments that were genotyped at 2533 informative molecular markers across 19Mb of the X chromosome. Spermatogenesis in sterile introgression genotypes proceeds through meiosis but does not complete spermatid individualization; this sterility phenotype appears distinct from that observed in F1 hybrid males. We have genetically dissected two small intervals to a dozen candidate hybrid sterility genes, and have identified and transgenically validated a new hybrid male sterility gene. Identification of multiple X-linked sterility genes will provide insight into the functions of these genes within species, the etiology of sterility, the evolutionary forces involved in interspecific divergence, and the overall molecular basis of the large X-effect in speciation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM7

How imprinted and imprinting incompatibilities impact introgression

Yaniv Brandvain [1,*]

[1]Department of Plant and Molecular Biology, University of Minnesota - Twin Cities, St Paul, United States

Abstract: Studies from plants and animals have shown that loci involved in genomic imprinting - the differential expression of an allele based on parental origin - can underlie hybrid incompatibilities. This involvement of genomic imprinting in reproductive isolation mechanisms could reflect either differential resolution of genomic conflict in diverging lineages, or disrupted co-adaptation of the imprinting machinery. Regardless, of the cause, the consequences of imprinting and/or imprinted loci for the patterns of introgression in hybridizing populations are unclear. I develop a model that tracks how asymmetric parent of origin incompatibilities influence patterns of introgression at linked and unlinked autosomes, sex-chromosomes and cytoplasm, and compare these results to expectations for biparental incompatibilities, maternal-effect incompatibilities and cytoplasmic incompatibilities. This work will guide empiricists in their interpretation of observed patterns of gene flow in taxa with incompatibilities known to be caused by imprinted alleles.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM10

Insights into the epigenetic basis of the Large X-Effect from studies in cat interspecific hybrids.

William Murphy [1,*], Wesley Brashear [1], Kevin Bredemeyer [1], Brian Davis [1], Gang Li [1], Elizabeth Heppenheimer [2], Christopher Seabury [1], Bridgett vonHoldt [2]

[1]Texas A&M University, College Station, [2]Princeton University, Princeton, United States

Abstract: Despite decades of mapping studies in model organisms, there has been limited progress made in identifying specific genomic mechanisms that would broadly explain the large role of the sex chromosomes in reproductive isolation across diverse organisms. Sex chromosomes have the lowest measures of contiguity and completeness in most draft genome sequences due to large numbers of moderate to large size repetitive elements, which can undermine efforts to characterize and resolve the genetic or epigenetic basis of sex-linked phenotypes. Here we describe results from a gene mapping study in an interspecies hybrid cat breed that exhibits hybrid male sterility, a cross between the domestic cat and Jungle cat. Gene mapping and targeting genome finishing were employed to fine map and resolve a structurally complex X chromosome locus that harbors a major effect hybrid sterility 'gene'. A combined analysis considering copy number variation, testis gene expression patterns, and methylation data provides novel insight into the molecular basis for hybrid sterility and the Large X-Effect in eutherian mammals.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

POB-189

The evolution of hybrid seed inviability in sympatric Mimulus species- a case for parental conflict?

Jenn Coughlan [1,*], John Willis [1]

[1]Biology, Duke University, Durham, United States

Abstract: Hybrid inviability is a common barrier to reproduction in plants. However, the genes causing hybrid inviability and the forces driving its evolution remain unclear. Here we explore a strong, asymmetric crossing barrier between naturally co-occuring diploid monkey-flowers- *M. guttatus* and *M. decorus.* Reciprocal crosses from across their ranges show that hybrid inviability is common and variable in both magnitude and direction of the crossing asymmetry, including in sympatric populations. In addition, asymmetries in reciprocal hybrid viability are associated with asymmetries in reciprocal hybrid seed size, suggesting both parent of origin effects and a role of developmental defects in endosperm as a cause of hybrid inviability. Intriguingly, we show that range-wide variation in reproductive isolation among populations of *M. decorus* with *M. guttatus* is associated with reproductive isolation between populations of *M. decorus,* wherein populations which differ in crossing patterns with *M. guttatus* exhibit extensive reproductive isolation, while those which show similar patterns of reproductive isolation with *M. guttatus* are cross compatible. Hybrid dysfunction and abnormal growth phenotypes persist in these crosses. Lastly, we use values of seed inviabilities in F1 backcrosses and F2s to determine that this incompatibility is caused by a nuclear incompatibility involving parent-of-origin effect alleles. We map the incompatibility using an F1 backcross, which allows us to determine the loci which are maternally or paternally inherited and contribute to seed inviability. We find that the hybrid seed incompatibility between *M. guttatus* and *M. decorus* is relatively simple, involving only a handful of loci in each species.

Expanded summary*: Biologists have long been astounded by the shear diversity of species in the world, and thus a main goal of evolutionary biology is to understand how species are formed. One of the most common and strong reproductive barriers is the formation of inviable hybrid zygotes. Despite its commonality, the genes which cause this incompatibility, as well as the selective forces driving the evolution of these genes remain unclear. One hypothesis for the evolution of hybrid inviability alleles is that of parental conflict, which posits that difference in resource allocation optima between mothers and fathers can lead to a co-evolutionary arms race in resource allocation alleles between males and females. Reproductive isolation then manifests when populations which are divergent in resource allocation alleles at two or more loci hybridize, and the resultant hybrids possess incompatible allelic combinations which result in inappropriate- and often lethal- growth phenotypes. While the parental conflict has gleaned much support theoretically, as well as in inter-ploidy crosses, much less is known about whether parental conflict can explain hybrid inviability in co-occuring, diploid plant species. Here I describe a strong, asymmetric crossing barrier between naturally co-occuring monkey flowers- *M. guttatus* and *M. decorus.* Reciprocal crosses from sixteen populations of *M. decorus* to four strains of *M. guttatus* show that hybrid inviability is both common and extremely variable in both magnitude and direction of the crossing asymmetry. These asymmetries in reciprocal hybrid viability are associated with asymmetries in reciprocal hybrid seed size, suggesting that hybrid inviability is a consequence of parent-of-origin effects, and that hybrid inviability is related to inappropriate growth, consistent with parental conflict. In addition, since populations of *M. decorus* show exceptional variation in their ability to cross to *M. guttatus,* we test whether hybrid seed inviability manifests between populations of *M. decorus,* and whether within-species incompatibilities are related to the strength/direction of between species incompatibilities. Indeed, we find that populatins of *M. decorus* which vary in their ability to cross to *M. guttatus* exhibit strong seed incompatibilities, while populations of *M. decorus* which show relatively similar patterns of crossing to *M. guttatus* are completely compatible. We then use values of seed inviabilities across F1 backcrosses and F2s to determine that this incompatibility is a nuclear-nuclear incompatibility between *M. guttatus* and *M. decorus* which involves alleles with parent-of-origin effects. Finally, we map this incompatibility using an F1 backcross design wherein an F1 individual was backcrossed to each parent reciprocally. This design allows us to determine the genetic loci which are maternally or paternally inherited and contribute to this incompatibility. We find that the hybrid seed incompatibility between *M. guttatus* and *M. decorus* is relatively simple, involving only a handful of loci in each species. Future work will aim to confirm that these loci are involved in asymmetric hybrid seed inviability and determine if the parent of origin effects are explained by genomic imprinting. Overall, we hope this work will contribute broadly to the study of the formation of species and barriers to reproduction upon secondary contact.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

POB-206

Venom as a “magic trait” - venom evolution in the spider genus Tetragnatha facilitates adaptive radiation and mate recognition

Michael Brewer [1,*], Zarif Hasan [1], Pamela Zobel-Thropp [2], Greta Binford [2], Rosemary Gillespie [3]

[1]Biology, East Carolina University, Greenville, [2]Biology, Lewis and Clark College, Portland, [3]ESPM, University of California, Berkeley, United States

Abstract: Venoms are considered a key innovation that allow ecological diversification, and often adaptive radiation, in many lineages, including snakes, anguimorph and iguanian lizards, and cone snails. Venoms are complex cocktails of bioactive compounds with high target specificity and serve several functions. Commonly, venom facilitates predation and defense but can function in intraspecific conflict and mate recognition.

Sexual dimorphisms have been observed in the venoms of several taxa, indicating a possible role in mate recognition. This posits a possible dual role of venoms in both ecological divergence and mate choice that could facilitate reproductive isolation of populations through local adaptation. Though venom sexual dimorphism can be due to differences in feeding behavior, evidence also suggests that venom may operate in intraspecific communication.

Our work on *Tetragnatha* venoms, (namely transcriptomics, proteomics, and stable isotope dietary analysis), shows most species display sexual differences in venom composition; mature males express large venom components that are virtually absent in females. Differential expression in venom gland tissues exists between sexes and habitats in a widespread, generalist North American species. Additionally, a Hawaiian clade has undergone an adaptive radiation accompanied by the evolution of venom peptides, including the novel venom gene families. Given the unusual mating behavior (cheliceral locking) in many tetragnathids, these dimorphic venom components may function in mate recognition while other components have evolved in response to dietary niche partitioning. We posit tetragnathid spider venom acts as a “magic trait” that facilitated local adaptation and at least one major adaptive radiation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

POB-207

Spiraling Complexity: A Test of the Snowball Effect in a Computational Model of RNA Folding

Ricardo Azevedo [1,*], Ata Kalirad [1]

[1]Biology and Biochemistry, University of Houston, Houston, United States

Abstract: Genetic incompatibilities can emerge as a by-product of genetic divergence. According to Dobzhansky and Muller, an allele that fixes in one population may be incompatible with an allele at a different locus in another population when the two alleles are brought together in hybrids. Orr showed that the number of Dobzhansky–Muller incompatibilities (DMIs) should accumulate faster than linearly—i.e., snowball—as two lineages diverge. Several studies have attempted to test the snowball effect using data from natural populations. One limitation of these studies is that they have focused on predictions of the Orr model but not on its underlying assumptions. Here we use a computational model of RNA folding to test both predictions and assumptions of the Orr model. Two populations are allowed to evolve in allopatry on a holey fitness landscape. We find that the number of inviable introgressions (an indicator for the number of DMIs) snowballs, but does so more slowly than expected. We show that this pattern is explained, in part, by the fact that DMIs can disappear after they have arisen, contrary to the assumptions of the Orr model. This occurs because DMIs become progressively more complex (i.e., involve alleles at more loci) as a result of later substitutions. We also find that most DMIs involve more than two loci—i.e., they are complex. Reproductive isolation does not snowball because DMIs do not act independently of each other. We conclude that the RNA model supports the central prediction of the Orr model that the number of DMIs snowballs, but challenges other predictions, as well as some of its underlying assumptions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

POB-197

Tetrahymena comparative genomics

Guangying Wang [1,2], Jie Xiong [2], Wentao Yang [1,2], Kai Chen [1,2], Chuanqi Jiang [1,2], Wei Miao [2,*]

[1]University of Chinese Academy of Sciences, Beijing, [2]Key Laboratory of Aquatic Biodiversity and Conservation, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China

Abstract: *Tetrahymena thermophila* is by far the most well-developed model organism for basic research among the ciliates. By choosing *Tetrahymena* species varying degrees of phylogenetic relatedness to *T. thermophila* with particular interesting ecological and morphological traits, we sequenced the MAC genomes of nine *Tetrahymena* species, including *T. thermophila*, *T. malaccensis*, *T. elliotti*, *T. pyriformis*, *T. vorax*, *T. borealis*, *T. sp*, *T. empidokyrea*, *T. shanghaiensis* and *T. paravorax*, to understand the speciation and adaptation evolution of *Tetrahymena.* Comparable good genome assembly to *T. thermophila* were gotten with almost all the N50 values large than 400Kb. Genome sizes of ten Tetrahyemna species range from 84.9 to 116.1Mb, and vary within 1.4 fold. With no doubt, all the genome of ten *Tetrahymenas* are AT rich, however, the GC content also vary from 20% (*T. shanghaiensis*) to 29% (*T. paravorax*). In general, all of ten *Tetrahymenas* have more than 20,000 protein coding genes, and the predicted protein coding genes vary within 1.3 fold range between ten species. Through constructing the *Tetrahymena* species tree using phylogenomic approach, the origin of *Tetrahymena* genus was estimated about 213-373Mya ago, suggesting the long evolution history of this genus. In summary, our data showed that genomes of ten *Tetrahymena* species are highly diverged, and the comparative genomic analysis will provided comprehensive understanding of the speciation and adaptation evolution of *Tetrahymena* species.

Expanded summary*: *Tetrahymena* is a genus of free-living ciliated protozoans that is widely distributed in freshwater environments around the world. *Tetrahymena thermophila*, best studied species in the genus, is well-established as a model eukaryote. By choosing *Tetrahymena* species varying degrees of phylogenetic relatedness to *T. thermophila* with particular interesting ecological and morphological traits, we sequenced the MAC genomes of nine *Tetrahymena* species, including *T. thermophila*, *T. malaccensis*, *T. elliotti*, *T. pyriformis*, *T. vorax*, *T. borealis*, *T. sp*, *T. empidokyrea*, *T. shanghaiensis* and *T. paravorax*, to understand the speciation and adaptation evolution of *Tetrahymena.*

Comparative genomic analysis shows there exists very high genetic divergence between these ten species, including varying genome size (ranging from 84.9 to 116.1Mb) and protein coding genes (1.3 fold range). Most importantly, we construct the *Tetrahymena* species tree using phylogenomic approach for the first time, and the origin of *Tetrahymena* genus was estimated about 213-373Mya ago, suggesting a long evolution history of this genus.

In summary, our data showed that genomes of ten *Tetrahymena* species are highly diverged, and the comparative genomic analysis will provided comprehensive understanding of the speciation and adaptation evolution of *Tetrahymena* species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

POB-198

Identifying and Analyzing Introgression Patterns in Capsella grandiflora

Krzysztof Stankiewicz [1,*], Tyler Kent [2], Stephen I. Wright [2], Yaniv Brandvain [1]

[1]College of Biological Sciences, University of Minnesota, St. Paul, United States, [2]Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Canada

Abstract: A transition in mating system, often by a loss of self-incompatibility, is a common mode of plant speciation. However, gene flow between selfing species and close outcrossing relatives is often possible as individuals from the two sister species can occasionally produce viable offspring leading to detectable traces of continuing admixture between these species. Given the changes associated with mating system shifts, including shifts in ecology, floral morphology and genetic load, there may be important effects of mating system transitions on the selective impact of introgression. A number of hypotheses have been proposed regarding determinants of local selection against admixture, including the influence of gene density, recombination rate, ecologically relevant QTL, and intrinsic incompatibilities.

To test these factors we developed a hidden Markov model-based method to identify regions of selfing genomes introgressed into an outcrossing relative based on rare genetic variants and haplotype data. We then applied the method to identify admixture in 182 individuals of the outcrossing *Capsella grandiflora* from its predominately selfing sister species *Capsella rubella*. Although these species diverged relatively recently (30-100 k.y.a.), *C. rubella* exhibits several selfing syndrome phenotypes which would likely be deleterious in *C. grandiflora*.

We combine our HMM output with previous data concerning local recombination rates, ecological QTL, and known incompatibilities between the *Capsella* species to test hypotheses involving introgression at these regions. We find no evidence that gene density, recombination rate, or incompatibilities influence the extent of local introgression. However, we find some evidence suggesting that ecologically relevant floral QTL may prevent introgression.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

POB-202

Inferring the onset of reproductive isolation from the genome-wide distribution of pairwise differences

Simon Aeschbacher [1,*], Konrad Lohse [2]

[1]Institute of Ecology and Evolution, University of Bern, Bern, Switzerland, [2]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

Abstract: Understanding the prevalence of different routes to speciation requires good estimates of the strength and timing of reproductive isolation. Here, we present a novel approach to date the onset of reproductive isolation from the genome-wide distribution of pairwise sequence differences among species. We obtain this distribution from analytic results for the distribution of pairwise coalescence times under a number of demographic scenarios that are of interest in the context of speciation. All else being equal, barriers to gene flow in genomic regions of low recombination reduce neutral divergence more strongly than barriers in regions of high recombination. Arranging the genomes of focal species pairs into recombination strata, we jointly fit our model to the distribution of pairwise differences across all strata to infer the strength and timing of effective gene flow. Applying our method to whole-genome sequence data from *Mimulus guttatus* and its selfing sister species *M. nasutus*, we confirm the onset of a strong reduction of effective gene flow about 500,000 years ago. We discuss the gain in temporal resolution of our new method compared to previous approaches, as well as its sensitivity to the choice of recombination strata.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

POB-204

Heterochromatin turnover among great apes: Species and gender differences

Monika Cechova [1,*], Robert S Harris [1], Kateryna Makova [1], Francesca Chiaromonte [1]

[1]Penn State, University Park, United States

Abstract: Heterochromatin comprises intriguing repeat-rich and gene-deprived genomic regions that play an essential role in the cell, both mechanistically (e.g., in centromeres and telomeres) and functionally. Such important processes as X-chromosome inactivation and tissue-specific gene repression are driven by heterochromatin. Moreover, even small segments where the heterochromatin is unbalanced can lead to mitotic failure and non-viable offspring in hybrids. Thus, rapid heterochromatin turnover can drive speciation. Indeed, heterochromatic repeats show remarkable diversity in composition and variation in repeat counts among species, populations, and individuals. Here, using short- and long-read resequencing data, we conduct the first detailed genome-wide investigation of heterochromatin turnover among great apes. We show that each species harbors only a handful of repeats, e.g., the (GGAAT)n pentamer in human and homologous 32-mers in chimpanzee, bonobo, and gorilla. Principal Component Analysis allowed us to successfully differentiate great ape species (but not sexes) based on abundances of the 10 most common repeats. We observed small interindividual variation in repeat counts in humans, whereas large interindividual variation characterizes gorillas and chimpanzee subspecies. We identified eight male-biased repeats that comprise up to 3.8% of the male genomes. To study repeat length distributions, we developed a novel algorithm, NoisyRepeatFinder that can analyze repeats in long and noisy PacBio reads. Lastly, we found that, for most linkage functions, unsupervised hierarchical clustering does not reproduce the expected species phylogeny, illustrating a remarkably high tempo of heterochromatin turnover in great apes that might have contributed to their speciation.

Expanded summary*: The heterochromatin has been for the first time described as early as in 1928, based on the contrasting compaction during interphase where it displayed as densely condensed chromatin. These intriguing inaccessible, repeat-rich and gene-deprived regions of the genome play an essential role in the cell, both mechanistically and functionally. Both centromeric and telomeric repeats are crucial during mitosis. Indeed, neocentromeres with paucity of heterochromatin domains have been shown to display defects during cell division. Important processes such as X-chromosome inactivation or tissue-specific gene silencing are also driven by heterochromatization. In order to prevent heterochromatin from uncontrolled spreading, other genomic elements - such as non-coding RNA - can be utilized in cell. Since our lab recently assembled the gorilla Y chromosome, we could trace the origin of many Y-repeats to the assembly steps where we introduced long PacBio data, recapitulating the notion that the heterochromatic repeats are often underexplored and underrepresented in the assemblies, especially those generated from short sequencing data. Indeed, it is estimated that as much as 5-10% of human genome is still unassembled, all of it consisting of heterochromatin. The situation is even worse for the sex-determining Y chromosome in human - more than half of its length is currently filled with stretches of unknown heterochromatic sequences. Repeats tend to collapse in the assemblies and their high similarity makes it very challenging to correctly assign their original genomic loci. Furthermore, heterochromatic repeats exhibit remarkable diversity among species, populations, and individuals. They are also extremely variable, for instance the length centromeric array of the X chromosome varies as much as ten-fold (0.5-5 Mbp) among individual humans.

Here, we survey heterochromatin turnover among five great apes using short- and long-read resequencing data. All five studied species are colonized by the (GGAAT)n repeat that is extremely abundant, present on all chromosomes including autosomes and whose function remains enigmatic, although earlier studies suggest that this motif could be overrepresented at the translocation breakpoints. We show that each species is populated by only a handful of repeats (a single abundant repeat for human and a maximum of ~15 repeats for chimpanzee). We focused on 10 most abundant repeats per species, to which we applied unsupervised and supervised clustering and classification algorithms (e.g. PCA and LDA) and demonstrated that, using only first three principal components, we could already separate the species into perfect clusters. We also show small variation in repeat counts in humans, but large variation in gorillas and chimpanzee

subspecies *Pan troglodytes verus.* Many of these repeat counts show high correlation with other (often similar) repeat counts, suggesting an underlying structure in the repeat count distribution.

In order to specifically study repeats located on the Y chromosomes, we focused our attention towards male-biased repeats (present consistently in higher abundances in males than in females). Male genomes of chimpanzee, bonobo, and gorilla (but not orangutan) are highly enriched in homologous 32-mers. We note that the limited Illumina read length enables one to study only repeat abundances, but not lengths. Therefore, we developed a new algorithm for the repeat length discovery. NoisyRepeatFinder accounts for the fact that long PacBio reads are populated by an extensive number of single-repeat insertions and uses built-in filters that remove homologous repeats mixed within the repeat of interest.

Proper understanding of repeat turnover will enable building better reference genomes, mitigating potentially severe effects of multimapping reads that obscure variant calling algorithms. Building models of repeat dynamics will help our understanding of individual variation and speciation. Finally, a better grasp of heterochromatin formation and spreading will clarify the regulation mechanism of many genes residing at or inside heterochromatin boundaries.

Taken together, we show heterochromatin as a dynamic and variable entity in great apes (especially gorillas) and provide a novel tool for the study of repeat length distribution from noisy PacBio data.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

POB-210

Sex determining genes have a greater effect on gene expression than sex chromosomes in the house fly, Musca domestica L.

Jae Hak Son [1,*], Tea Kohlbrenner [2], Svenia Heinze [2], Daniel Bopp [2], Richard Meisel [1]

[1]Department of Biology and Biochemistry, University of Houston, Houston, United States, [2]Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland

Abstract: Sex determination (SD) evolves fast. Sexually antagonistic selection at an autosomal locus can induce the invasion of a linked new SD gene, driving the evolutionary turnover of sex chromosomes. However, it has not been tested whether a new SD gene, rather than linked alleles, confers phenotypic effects that can be acted upon by natural selection. We used house fly to test this hypothesis. The house fly male-determining (M) gene is a negative regulator of *Md-tra*. A chromosome carrying M is a neo-Y chromosome that is inherited by males, which should favor the fixation of male-beneficial alleles on the neo-Y. We injected dsRNA to knockdown *Md-tra* and create sex-reversed males that do not carry M or Y-linked alleles, and we compared the sex-reversed (XX) males with normal (XX) females and (XY) males. We used RNA-seq to determine the effects of SD genes and Y-linked alleles on gene expression. The expression profiles of sex-reversed (XX) males were similar to normal (XY) males. In contrast, about two thousand genes are differentially expressed between sex-reversed (XX) males and normal (XX) females, a similar magnitude as between normal (XY) males and (XX) females. We therefore concluded that SD genes (M and *Md-tra*) have greater effects on gene expression than X- and Y-linked genes in house fly. This suggests that the fitness effects of SD genes themselves should be considered in models of the evolutionary turnover of SD and sex chromosomes.

Expanded summary*: Sex determination (SD) mechanisms evolve rapidly. It is of great interest to identify the evolutionary forces that drive rapid change of SD. Several models have been developed to explain the divergence of SD mechanisms: random genetic drift, pleiotropic selection, sex-ratio selection, and sexually antagonistic selection. The sexually antagonistic selection hypothesis has been often used to explain the evolutionary turnover of SD and sex chromosomes. A sexually antagonistic locus on autosomes is under opposing selection pressures in males and females. The sexual conflicts can be resolved if a new SD locus arises and is linked to sexually antagonistic alleles. This is because, with linkage to the new SD locus, the sex-beneficial allele will be inherited only in the sex in which it is beneficial. In contrast, direct selection on phenotypic effects of the SD gene itself has received considerably less attention as a mechanism for driving evolutionary turnover in SD pathways.

The house fly is a good model system to compare the effects of SD genes and linked alleles on the evolution of SD pathways because it has an especially labile SD system. The male-determining (M) gene can be located on multiple different chromosomes, and each M-bearing chromosome is a neo-Y chromosome. The Y-linked alleles are inherited through males, which should favor the accumulation of male-beneficial mutations. Because M is a negative regulator of *Md-tra*, knocking down *Md-tra* with RNAi has a similar effect as M, causing XX embryos to develop into fertile males that lack Y-linked alleles. We took advantage of this flexibility to test whether SD genes, rather than Y-linked alleles, confer phenotypic effects that could be acted upon by natural selection.

We used RNAi knockdown of *Md-tra* to produce sex-reversed (XX) males, and we used RNA-seq in normal (XY) males, (XX) females, and sex-reversed (XX) males to evaluate the effect of SD genes and Y-linked alleles on gene expression. The magnitude of differential gene expression between sex-reversed (XX) males and normal (XX) females is similar to that between normal (XY) males and (XX) females. This indicates that SD genes have greater effects on gene expression than X- and Y-linked genes in house fly. This result suggests that phenotypic effects of SD genes themselves should be considered as potential drivers of the evolutionary turnover of sex chromosomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM12

The Spectrum of Epistasis and its Consequences for Hybrid Fitness and Speciation

Andrius Jonas Dagilis [1,*], Daniel Bolnick [1]

[1]Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: Hybrids often exhibit either vigor or inviability. We use an interaction network of yeast consisting of 20 million interactions between 6018 genes to investigate what kinds of genes cause reproductive isolation and heterosis. Using the joint distribution of epistatic interactions and the fitness of genes that underlie them as a guide to generate a Bateson Dobzhansky Muller incompatibility model, we develop a common explanation for heterosis and reproductive isolation. We find that very few interactions with large effect on hybrid fitness are likely to occur over the course of speciation. These results suggest that many interactions of small effect may be the primary cause of reproductive isolation. We further find that different genes are responsible for reproductive isolation early vs late in divergence. Our study outlines the potential to perform predictive speciation genomics using interaction network data.

Expanded summary*: When studying speciation, it is difficult to predict which genes will be responsible for reproductive isolation between two populations. This is in part because the epistatic interactions that underlie Bateson-Dobzhansky-Muller (BDM) incompatibilities are poorly studied. For example, there are nearly 5000 potential interactions between alleles in an organism with only 100 biallelic genes. Thus, measurements of strength of epistasis to date have been restricted to a relatively small number of epistatic interactions. A recent study by Costanzo et al () provides an unprecedented dataset for studying the epistatic interactions that can drive speciation. By measuring the fitness of more than 20 million double knockout *S. cerevisiae* strains, they were able to identify the strength of epistatic interactions between 10,000 single knockout strains. While knockouts do not have fitness effects that are representative of the average mutation, this dataset represents the best-measured interaction network to date.

We reanalyzed the Costanzo et al. data from the perspective of speciation genetics and asked what kinds of interactions drive reproductive isolation. We first calculated the probability that any single knockout would fix in a population with no other mutations. We then asked what interactions were likely to occur when crossing two populations with different fixed mutations. We found that there was a significant positive relationship between the strength of an epistatic interaction and the probability that such an interaction would be found in hybrids. In fact, many epistatic interactions between the genes that were likely to fix early were positive, suggesting that hybrid vigor may frequently occur during the early stages of divergence. We then simulated sequential fixation of substitutions in two different populations, accounting for epistasis with previously fixed alleles. We found that hybrids between simulated populations show a decline in relative fitness as the populations diverge, but can also exhibit hybrid vigor for long timespans. Furthermore, the drivers of reproductive isolation vary between early and late divergence. GO enrichment analyses confirm that different pathways are likely to contribute to divergence at early vs late stages of speciation. In general, we find that very few interactions of large effect are likely to cause reproductive isolation, and the decrease in hybrid fitness occurs due to the cumulative effects of many interactions of weak effect.

Using the distribution of epistatic interactions as a guideline, we develop a mathematical model of BDM incompatibilities that involves both positive and negative epistasis. We find that relative hybrid fitness depends on two general factors – the addition of new epistatic interactions in the hybrid, and the dilution of interactions between co-adapted sets of genes from the parental populations. The model is able to predict the conditions under which hybrid vigor is expected as well as when the "snowball effect" predicted by other speciation models should exist. In this way, it unites the often-disparate fields of theory concerning heterosis and speciation. The model suggest that positive epistasis may also contribute to some of the most prominent patterns in speciation genetics, such as Haldane's Rule or Darwin's Corollary. Combined with our simulation results, the model suggests that different processes may underlie reproductive isolation at different stages of speciation. Finally, the study presents a testable prediction of which genes are likely to contribute to speciation in lab strains of *S. cerevisiae.*

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM13

The molecular genetic basis of recombination variation in adaptively diverging stickleback fish

Vrinda Venu [1,*], Enni Harjunmaa [1], Felicity Jones [1]

[1]FML, Max Planck Society, Tuebingen, Germany

Abstract: Meiotic recombination is one of the major molecular mechanisms generating genetic diversity. By making and breaking new allelic combinations it can directly influence the patterns and efficacy of natural selection. Using the three spine stickleback fish, an evolutionary model organism, we are investigating the role and influence of meiotic recombination on the distribution, segregation and molecular basis of marine and freshwater adaptive divergence. We are using both whole genome sequencing and ChIP-Sequencing to empirically identify and quantify meiotic cross-overs for multiple individual fish. We start by constructing high-resolution individualized maps of meiotic recombination events by whole genome sequencing of large nuclear families (2 parents &~100offspring).This is complemented with ChIP-Sequencing of candidate meiotic proteins to label genomic cross-over locations with high resolution. Using these approaches with multiple individuals and families we aim to quantify variation in recombination hot and cold spots across the genome, and compare recombination maps among different individuals, sexes and adaptively diverging stickleback species. Preliminary results based on multiple individual high-resolution genome-wide maps of recombination show striking differences in cross-over patterns between males and females, and suggest the presence of both recombination “hot-” and “cold-spots”. Adaptive loci tend to fall in regions of low recombination suggesting maintenance of linkage among adaptive alleles is important during adaptive divergence with ongoing gene-flow. Ongoing work including analysis of DNA motifs, forward genetic mapping of recombination modifiers and tests for molecular signatures of selection aim to understand how this key biological process facilitates or constrains adaptive evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM2

Gene interactions and coadaptation in the evolution of genomic imprinting

Jason Wolf [1,*]

[1]Milner Centre for Evolution, University of Bath, Bath, United Kingdom

Abstract: For genes showing genomic imprinting, the expression of the copy inherited from mothers differs from the expression of the copy inherited from fathers. Most often, this differential expression arises from epigenetic silencing of one copy. When there is allelic variation at an imprinted locus, this parent-of-origin dependent pattern of expression leads to a different genotype-phenotype relationship for allelic variation inherited from mothers than from fathers. Consequently, genomic imprinting can also modify the contribution of allelic variation to patterns of quantitative genetic variation and the resemblance of relatives. I will discuss how this phenomenon can potentially favour the evolution of genomic imprinting, especially in cases where there are gene interactions, including both interactions between loci within a genome as well as interactions between the genotypes of individuals in social interactions. I will focus on scenarios where imprinting is favoured because it modifies patterns of allelic interactions and thereby facilitates evolutionary coadaptation. I will discuss the consequences of this evolutionary process for patterns of imprinting across loci within a genome and across orthologs in different genomes. Finally, I will discuss the consequences of evolutionary differences in imprinting for population differentiation and speciation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

POB-208

The evolution of aminoacyl-tRNA synthetases in chromerids

Abdoallah Sharaf [1,*], Kateřina Jiroutová [1], Miroslav obornik [1]

[1]Institute of Parasitology, Biology Centre ASCR, České Budějovice, Czech Republic

Abstract: Aminoacyl-tRNA synthetases (aaRS) are enzymes catalyzing the ligation of tRNAs to their cognate amino acids. There are aaRSs specific for each of the 20 standard amino acids. These enzymes are divided into two classes, class I and class II, which are unrelated in both sequence and structure. The aaRSs, are particularly prone to functioning in multiple sub-cellular compartments. Regard to a phenomenon known as dual targeting. We searched the genomes of the *Chromera velia* and *Vitrella brassicaformis* for aaRSs genes. A phylogenetic analyses of all available 21 aaRSs sequences were performed using maximum likelihood method and Bayesian inference. Computer predictions of the intracellular location of the identified enzymes were performed to test the multiple targeting hypothesis. Fifty genes encoding aaRS were identified in *C. velia*, while only 38 aaRSs were found in *V. brassicaformis*. Only α subunit of pheRS was found in *V. brassicaformis* in three copies; 71% of aaRSs were encoded by two genes and 23% of aaRSs are single copy in the algal genome. In contrast, *C. velia* contains three copies of particular aaRS in 38% cases, 42% of aaRSs were shown to be present in two copies and 14% of the genes are single copy. Most of the molecular phylogenies of aaRSs showed that for each aaRS the evolutionary pattern is different and eukaryotic genes are usually retained. Targeting predictions showed that particular enzymes are not often used in compartments they originate from.

Expanded summary*: Aminoacyl-tRNA synthetase (aaRS) is an enzyme that catalyzes the ligation of tRNAs to their cognate amino acids. There are aaRSs specific for each of the 20 standard amino acids. The aminoacyl-tRNA synthetases (aaRSs), are particularly prone to functioning in multiple sub-cellular compartments. Regard to a phenomenon known as dual targeting.We searched the genomes of the *Chromera velia* and *Vitrella brassicaformis* for aaRSs genes. Localizations prediction of the identified genes were performed to test its multiple targeting hypothesis. In order to examine the complex evolutionary path of the aminoacyl-tRNA synthetases, A phylogenetic trees of a universal Datasets of all available 21 aminoacyl-tRNA synthetases sequences for Archaea, Bacteria, Cyanobacteria, Opisthokonts, Amoebozoa, Archaeplastida, secondary algae and Excavata groups representatives were computed using gamma-corrected LG substitution matrix.

Fifty aaRS loci were identified in *Chromera velia* and 38 aaRS loci in *Vitrella brassicaformis*. Although this is substantially fewer than the number required to provide each translationally active compartment with its own unique protein (roughly 60 for each), aaRSs are highly conserved and it is unlikely that any were missed by our search. Multiple targeting was anticipated for one or more members of a given aaRS type unless there were sufficient numbers of genes to provide unique proteins for each sub-cellular compartment. Only one aaRS type in *V. brassicaformis* was represented by three loci (pheRS alpha subunit) and the rest aaRS types were represented by two loci (71%) and one loci (23%), while 38% of aaRS types in *C. velia* were represented by three loci, 42% were represented by two loci and 14% by one locus. Interestingly, valRS was represented by five loci. Dual targeting of several identified aaRSs loci in both *C.velia* and *V. brassicaformis* was confirmed using the localization prediction. Most of the molecular phylogenies of aaRSs showed the non-canonical pattern, Moreover, the plastid-targeted aaRS branched with cyanobacteria. Which is clear indications that endosymbiotic horizontal transfer has occurred during the evolutionary history of aaRSs.

The results of this study provide the first report of aaRS, its multiple targeting and evolution in Chromerida, As a first step toward a more nuanced understanding of protein targeting in these complex algae. Finally, we introduce a robust evidence of endosymbiotic horizontal gene transfer in Chromerida.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM11

The origin and evolution of genomic imprinting in the placenta of therian mammals from marsupial-eutherian comparative genomic analyses

Xu Wang [1,*], Kory Douglas [2], Paul Samollow [2], Andrew Clark [1]

[1]Cornell University, Ithaca, [2]Texas A&M University, College Station, United States

Abstract: In animals, genomic imprinting arose in the ancestor of therian mammals, which coincide with the emergence of viviparity, a key evolutionary innovation that improves reproductive success through placental *in utero* development. To investigate the origin and evolution of imprinting, we profiled genome-wide allele-specific expression, histone modifications and DNA methylation in the placenta of several mammals, and compared the imprinting profile between eutherians and a marsupial species, the gray short-tailed opossum. Although the imprinting status is conserved for the majority of eutherian imprinted genes, only less than 10% of their opossum orthologs are imprinted, reflecting striking evolutionary fluidity. Opossum-specific imprinted genes are expressed in early conceptus and reproductive systems with important roles in embryonic and placental development. Three novel imprinted genes in opossum are derived from marsupial-lineage specific duplications, suggesting a reduction of expression on the imprinted copy is favored after duplication and stable imprinting may arise due to neo-functionalization in placenta. We estimate that opossums imprint only 30-40 genes, or about 1/5 the number imprinted by eutherian mammals. The smaller number and non-overlapping nature of imprinted genes could be due to the primitive placentation and shorter gestation time in marsupials compared to eutherians. Our study will shed light on the origin and evolution of imprinting in mammals.

Expanded summary*: Genomic imprinting is an epigenetic phenomenon in which the expression of a subset of genes is in a parent-of-origin dependent manner. Genomic imprinting represents loss of diploidy in gene expression, resulting in effectively hemizygous state with vulnerability to deleterious mutations. The origin of genomic imprinting and how imprinting status is fixed in the population has become an evolutionary puzzle. The most popular model is the parental genetic conflict hypothesis: paternal alleles have maximal fitness if their fetus grows as large as possible at the expense of the maternal resources; maternal alleles have improved fitness if the mother distributes her resources equally to multiple litters to increase the genetic diversity. The parental conflict hypothesis is well accepted by the imprinting field and it does an excellent job explaining the origin of several famous imprinted genes: paternally expressed imprinted genes tend to be growth factors and maternally expressed imprinted genes are growth inhibitors. One limitation of this hypothesis is that only 9 of ~200 known imprinted genes have a growth phenotype and there is one exception (a maternally expressed growth factor). Recently, a number of other genetic conflict and non-conflict hypotheses have been developed.
The lack of imprinting in monotremes suggests that it arose in the ancestor of therian mammals. To date, ~180 imprinted genes have been discovered in mouse and ~100 in human. Not much is known outside the two species. Another hypothesis on the origin of genomic imprinting is that imprinting co-evolved with mammalian placenta and in utero development of the fetus, based on the observation that most imprinted genes are imprinted in fetus and the placenta. To test this hypothesis and investigate the origin and evolution of imprinting in therian mammals, comparative genomic analyses between eutherian mammals and marsupials is extremely informative. I quantified genome-wide parent-of-origin effect in gene expression by profiling allelic expression using RNA-seq in reciprocal hybrid F1s of several mammalian species, including inbred mouse strains, horse and donkey, cow and bison, two opossum strains and two inbred chicken breeds. This study provides the first comprehensive catalog of parent-of-origin expression status in vertebrates, opens the door to mechanistic analysis of marsupial-specific imprinted genes and sheds light on both the regulation and evolution of genomic imprinting in mammals.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM5

Biological species are universal across Life's domains

Louis-Marie Bobay [1,*], Howard Ochman [1]

[1]Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: Delineation of species is fundamental to organizing and understanding biological diversity. The most widely applied criterion for distinguishing species is the Biological Species Concept (BSC), which defines species as groups of interbreeding individuals that remain reproductively isolated. The BSC has broad appeal; however, many organisms, most notably asexual lineages, cannot be classified according to the BSC. Despite their exclusively asexual mode of reproduction, Bacteria and Archaea can transfer and exchange genes though homologous recombination. Here we show that barriers to homologous gene exchange define biological species in prokaryotes with the same efficacy as in sexual eukaryotes. By analyzing the impact of recombination on the polymorphisms in thousands of genome sequences, we find that over half of named bacterial species undergo continuous recombination among sequenced constituents, indicative of true biological species. However, nearly a quarter of named bacterial species show sharp discontinuities and comprise multiple biological species. These interruptions of gene flow are not a simple function of genome identity, indicating that bacterial speciation does not uniformly proceed by the gradual divergence of genome sequences. The same genomic approach based on recombinant polymorphisms retrieves known species boundaries in sexually reproducing eukaryotes. Thus, a single biological species definition based on gene flow is applicable to all cellular lifeforms.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM6

Human X and Y chromosome co-evolution, ampliconic gene evolution, selective sweeps and speciation

Elise Lucotte [1], Kasper Munch [1], Moises Coll [1], Marlene Dalgaard [2], Kristian Almstrup [2], Mikkel Heide Schierup [1,*]

[1]Bioinformatics Research Centre, Aarhus University, Aarhus, [2]Growth and reproduction, Rigshospitalet, Copenhagen, Denmark

Abstract: The X chromosome is disproportionally involved in speciation in humans and other great apes. We recently reported that the X chromosome has been the target of independent very strong selective sweeps in several great apes species targeting overlapping regions. These regions associate with the location of multi-copy, testis-expressed genes (so-called ampliconic genes) and also with genomic deserts of Neanderthal introgression into humans from interbreeding around 50,000 years ago. This suggests that these regions contain reproductive incompatibilities between human and Neanderthal, possibly due to the ampliconic genes. We speculated that competition between X and Y in male meioses, i.e. meiotic drive, by these ampliconic genes and their non homologous counterparts on the Y chromosome is responsible for these sweeps, and that such drive may be a major contributor to speciation. We present results on the variation in ampliconic gene copy number within and among human populations based on a new mapping approach of short read sequences from the Simons genome diversity and the Danish pangenome projects. We report extensive variation in ampliconic gene number for 7 X-linked and 7 Y-linked ampliconic regions and find that this variation is geographically structured around the globe. For the Y-chromosome, many duplications and deletions of ampliconic genes occur recurrently among different haplogroups. We relate this variation to our inference of very strong X-linked selective sweeps targeting specific human populations in order to identify potential drivers. Finally, we present preliminary results on ampliconic gene expression through male meiosis studied from micro-dissection of testes, and how this expression relates to the copy number of ampliconic genes and the ratio of X and Y chromosomes in spermatozoa.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

POB-184

Genome-wide characterization of population-specific mutation fitness effects

Alec Coffman [1], Alyssa Fortier [1], Jose Enrique Leon Burguete [2], Aaron Ragsdale [3], Ryan Gutenkunst [1,*]

[1]Molecular and Cellular Biology, University of Arizona, Tucson, United States, [2]Genomic Sciences, Universidad Nacional Autónoma de México, Cuernavaca, Mexico, [3]Applied Mathematics, University of Arizona, Tucson, United States

Abstract: Divergent selection, in which the same allele has different effects on fitness in different populations, drives environmental speciation. Much is known about patterns of genetic variation near loci with given divergent selection coefficients, but little is known about overall genomic patterns of divergent selection. To fill this gap, we developed a framework for inferring the joint distribution of fitness effects (DFE) between pairs of populations, based on the joint allele frequency spectrum. We applied this framework to African and European populations of Drosophila melanogaster, first estimating demographic history and then the joint DFE. As expected, we found that genome-wide nonsynonymous mutation fitness effects were highly correlated between these two populations. We then considered functional subsets of genes, however, seeking to identify functions on which selection is particularly divergent.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM14

Genomic Imprinting and Speciation in Mammals

Jeffrey Good [1,*], Thomas Brekke [1 2], Shane Campell-Staton [1 3], Zachary Cheviron [1]

[1]Biological Sciences, University Of Montana, Missoula, United States, [2]Biological Sciences, Bangor University, Bangor, United Kingdom, [3]Integrative Biology, University of Illinois, Champaign-Urbana, Champaign, United States

Abstract: Extreme hybrid growth is common in mammals, indicating that disruption of early development may play an important role in mammalian speciation. Disruption of genomic imprinting, the parent-specific epigenetic silencing (imprinting) of one allele, in the placenta has been hypothesized to be the predominant cause of abnormal hybrid growth. We have combined transcriptomic and quantitative genetic experiments to dissect the regulatory underpinnings of extreme parent-of-origin hybrid overgrowth between two species of dwarf hamsters. First, we tested for disrupted placental gene expression in overgrown F1 hybrid placentas. We observed extensive transgressive bi-allelic expression at several growth-related genes consistent with the disruption of paternal imprinting in overgrown hybrid placentas. Next, we used a systems genetics approach to examine the genetic architecture of hybrid placental overgrowth, gene expression, and imprinting in a hybrid mapping panel. The X chromosome emerged as the major maternal factor explaining hybrid placental overgrowth and the associated disruption of placental gene expression. Genome-wide expression analyses on backcross placentas revealed correlated networks of imprinted and non-imprinted autosomal genes that were misexpressed dependent on X-linked epistatic interactions. However, imprinted expression of the X chromosome appeared unperturbed in the same crosses. Collectively, our results suggest that the X chromosome may play an important role in the evolution of imprinted gene expression networks, and that the disruption of such networks contributes to the evolution of extreme hybrid growth in mammals.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Genomic mechanisms of speciation

OTH-GM3

Hybrid seed failure among wild tomato lineages: perturbed expression and imprinting in the endosperm

Morgane Roth [1,*], Ana Marcela Florez-Rueda [2], Margot Paris [3], Thomas Städler [1]

[1]Institute of Integrative Biology, ETHZ, [2]Department of Plant and Microbial Biology, University of Zürich, Zürich,

[3]Department of Biology, University of Fribourg, Fribourg, Switzerland

Abstract: Endosperm misdevelopment leading to hybrid seed failure (HSF) is a cause of reproductive isolation in angiosperms. This phenotype is found in interploidy and homoploid crosses between closely related species. Parent-of-origin expression (genomic imprinting) is thought to be necessary for normal endosperm development. Under this premise, a mismatch in imprinting between related species might cause HSF. Yet published studies have not addressed the potential correlation between imprinting divergence and HSF. We studied these issues using intra- and interspecific crosses of three diploid, self-incompatible tomato lineages. *Solanum arcanum* var. marañón, *S. chilense*, and *S. peruvianum* were chosen for their recent divergence and partial to near-complete reciprocal HSF. Based on transcriptome sequencing of laser-extracted endosperm, we quantified the extent of shared imprinting among the three reciprocal intraspecific crosses. Candidate imprinted genes have higher mean expression levels compared to non-imprinted genes, implying that specific mechanisms contribute to their apparent upregulation. In the two hybrid crosses with strong HSF, maternal proportions are increased and imprinting is perturbed. Such marked changes were not detected in the hybrid cross resulting in partial HSF, suggesting that the extent of misexpression covaries with the degree of HSF. Future inclusion of self-compatible tomato species may reveal whether imprinting dynamics is driven by parental conflict.

Expanded summary*: My goal is to better understand the genetic mechanisms of speciation in plants. Especially, what drives the establishment of reproductive barriers? The tomato clade (*Solanum* section *Lycopersicon*), encompassing at least 13 closely-related species with partial-to-complete hybrid seed failure phenotypes and variable outcrossing rates, is an outstanding system to study this question.

I currently focus on the phenotypic and molecular correlates of hybrid seed failure (HSF) in wild tomatoes. We found histological evidence that HSF is caused by impaired endosperm development at early stages. This phenotype is thought to be impacted, at least partially, by differences in patterns of genomic imprinting between parental species. The parental conflict theory posits that imprinting is driven by divergent parental interests in outcrossing species. Imprinting status may evolve rapidly, thus creating incompatibilities between nascent species. However, only few plant systems allow for testing imprinting mismatch in closely-related species with varying levels of HSF and divergent mating systems.

Among wild tomato lineages, *S. chilense* (C) and *S. peruvianum* (P) are genetically very close yet their reciprocal hybrid seeds largely inviable. *Solanum arcanum* var. marañón (A) is more diverged but genetically equidistant from C and P. Surprisingly, AP hybrid seeds fail to develop while an intermediate proportion of AC hybrid seeds appears to be viable. I have dissected these crossing disparities with transcriptome data obtained from laser-extracted endosperm. My work (i) quantifies the extent and divergence of genomic imprinting in normally developing crosses, and (ii) compares these data to the (partly perturbed) parental expression proportions observed in hybrid endosperm.

Comparing parent-specific allele expression, I showed that many genes imprinted in normally developing seeds lose their parental bias in strongly abortive seeds. This suggests that imprinting (at least at some genes) is not only functionally relevant for seed development, but also contributes to reproductive isolation between wild tomato lineages.

My project assesses the potential significance of imprinting at early stages of the speciation process. In the near future, we will test whether parental conflict drives the evolution of imprinted genes in *Solanum* by quantifying sequence polymorphism among many wild tomato species; this has not yet been addressed in any other plant system.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-321

Influenza A virus avian-derived PB1 gene has evolved to match its codon usage to interferon-altered human tRNA pools

Bartram Smith*, Guifang Chen, Claus Wilke, Robert Krug

Abstract: Influenza A viruses cause an annual highly contagious respiratory disease in humans, and are responsible for periodic human pandemics that have high mortality rates. Pandemic influenza A viruses can result from reassortment of one or more gene segments between a human and an avian virus. These avian virus gene segments need to adapt to humans post introduction. The role of synonymous mutations in this adaptation is not known. Here we focus on the human adaptation of the synonymous codons of the avian virus PB1 gene of the 1968 H3N2 pandemic virus. We generate recombinant H3N2 viruses differing only in codon usage of PB1 mRNA, and demonstrate that the codon usage of recent virus isolates enhances replication in interferon-treated human cells rather than in untreated cells, thereby partially alleviating the interferon-induced antiviral state. High-throughput sequencing of tRNA pools explains this virus phenotype: the levels of some tRNAs differ between interferon-treated and untreated human cells; and the codon usage of H3N2 PB1 mRNA has been evolving over time to match the tRNA pools in interferon-treated human cells. Consequently, our results identify a previously unknown mechanism by which influenza A virus counteracts the host interferon-induced antiviral response and highlight the important role of tRNA pools in the regulation of gene expression.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-319

Functional mutations in spike glycoprotein of Zaire ebolavirus associated with an increase in infection efficiency

Mahoko Ueda [1], Yohei Kurosaki [2], Taisuke Izumi [3], Yusuke Nakano [3], Olamide Oloniniyi [2], Jiro Yasuda [2], Yoshio Koyanagi [3], Kei Sato [3], So Nakagawa [4,*]

[1]Micro/Nano Technology Center, Tokai University, Hiratsuka, [2]Department of Emerging Infectious Diseases, Institute of Tropical Medicine (NEKKEN), Nagasaki University, Nagasaki, [3]Laboratory of Systems Virology, Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto, [4]Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Japan

Abstract: Ebola virus (EBOV) is extremely virulent, and its glycoprotein is necessary for viral entry. EBOV may adapt to its new host humans during outbreaks by acquiring mutations especially in glycoprotein, which allows EBOV to spread more efficiently. To identify these evolutionary selected mutations and examine their effects on viral infectivity, we adopted experimental–phylogenetic–structural interdisciplinary approaches. In evolutionary analysis of all available *Zaire ebolavirus* glycoprotein sequences, we detected two codon sites under positive selection, which are located near/within the region critical for the host-viral membrane fusion, namely alanine-to-valine and threonine-to-isoleucine mutations at 82 (A82V) and 544 (T544I), respectively. The fine-scale transmission dynamics of EBOV Makona variants that caused the 2014-2015 outbreak revealed that A82V mutant was fixed in the population while T544I was not. Further pseudotype assays for the Makona glycoprotein demonstrated that the A82V mutation caused a small increase in viral infectivity compared with the T544I mutation. These findings suggest that mutation fixation in EBOV glycoprotein may be associated with their increased infectivity levels; the mutant with a moderate increase in infectivity will fix. Our findings demonstrated that a driving force for Ebola virus evolution via glycoprotein may be a balance between costs and benefits of its virulence.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-325

Evolution of hosts, parasites and their microbiomes

Nolwenn Dheilly [1,*], Megan Hahn [1]

[1]SoMAS, Stony brook University, Stony Brook, United States

Abstract: It is becoming evident that microbes can affect the outcome of interactions between metazoan parasites and their hosts: Microbes modulate the maturation and efficiency of the immune system, and in some cases participate directly in defense mechanisms. More poorly recognized, however, is that parasites can also interact with microbes, and the outcomes of these interactions for the diseases they cause is unclear. The role of microbes in host-parasite co-evolution remains to be determined and could provide new clues for targeted intervention reducing parasite development, transmission and pathogenicity. We have identified the tapeworm *Schistocephalus solidus*, a common parasite of Threespine Stickleback fish, *Gasterosteus aculeatus*, as a very promising tractable model that allows experimental infections as well as field studies. We have characterized the bacterial microbiome of *S. solidus* and showed similarities with the fish microbiome, as well as strong selection for community composition. In addition, comparative analyses revealed that infection by the cestode parasite correlates with a disruption of the stomach and intestine bacterial microbiomes of sticklebacks characterized by an enrichment in key bacterial genus. Virus particles have also been purified from the intestine of healthy and parasitized sticklebacks and from *S. solidus* and observed on epifluorescence microscope. The viral microbiomes (RNA and DNA viruses) are currently been sequenced. Our results suggest a deterministic selection for community membership by the parasite that is limited by the composition of the host microbiome. Our next steps are (i) to determine whether the parasite is locally adapted to its host microbiome and (ii) to develop tools to manipulate the parasite microbiome and test its role in the disease.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-323

Extreme QTL and pooled sequencing for determining the genetic basis of host specificity in schistosome parasites

Frédéric D. Chevalier [1,*], Winka Le Clec'h [1], Benjamin Gourbal [2], Guillaume Mitta [2], Tim J. C. Anderson

[1]Genetics, Texas Biomedical Research Institute, San Antonio, United States, [2]Université de Perpignan Via Domitia, Perpignan, France

Abstract: Interactions between parasitic trematodes and their aquatic snail hosts provide a classical example of gene-for-gene co-evolution. Trematode infections typically sterilize snails, leading to selection of costly defense mechanisms, while parasites evolve to circumvent these defenses. We used a genetic approach to identify the parasite genes involved in overcoming snail defenses in the *Biomphalaria glabrata* (snail) - *Schistosoma mansoni* (parasite) system. We performed genetic crosses between two schistosome populations (SmBRE and SmLE) with distinctive patterns of host specificity: while both parasite populations infect BgBRE snails, only one (SmLE) can infect a second snail population (BgBS90). The F1 parasite progeny from our crosses were unable to infect BgBS90, while ability to infect BgBS90 snails was recovered in some F2 progeny. To identify the genome regions involved in snail specificity we used the extreme QTL (X-QTL) approach, developed by yeast and malaria researchers. In this method, pooled F2 progeny are selected for the trait of interest (i.e. ability to infect different snail populations), and then pools of selected or unselected progeny are quantitatively genotyped to measure allele frequencies genome-wide. We compared exome sequences of F2 parasite pools before and after infection of the two snail lines. Two genome regions (on chromosome 2 and 3) showed dramatic allele frequency distortion in parasites infecting BgBS90, and clearly underlie host specificity. We will now identify the specific genes involved in host specificity using RNAi approaches. Our long-term aim is to identify interacting genes in both parasite and snail to understand host-parasite evolution at the molecular level.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-322

NS2b3 Cleavage of STING is a Likely Barrier to Dengue Transmission Between Primate Hosts

Alex Stabell [1,*], Sara Sawyer [2]

[1]Department of Molecular Cell and Developmental Biology, [2]Biofronteirs Institute, University of Colorado Boulder, Boulder, United States

Abstract: Dengue virus circulates worldwide as four *endemic* serotypes (DENV1-4) and infects about 100 million people each year. Each of these serotypes likely originated in Southeast Asia as the result of an independent zoonotic transmission of a monkey-adapted (sylvatic) dengue to humans, followed by a sustained transmission cycle in the human population. Though the virus originated in Southeast Asia, it has spread globally, establishing sylvatic strains in new primate hosts and spreading to humans as well. We show that selective forces have driven differences in the primate innate immune response to dengue infection. Specifically, there is strong evidence for positive selection in the STimulator of Interferon Genes (STING). STING is typically cleaved by the dengue NS2b3 protease during viral infection. Here, we show that STING is highly variable in primates, causing differential susceptibility to cleavage by the dengue protease. In addition, we have used within population SNP data to detect a more recent signature of selection in chimpanzees, where the virus has more recently spread. Interestingly, sites detected to be under positive selection in chimpanzees have rendered this species STING cleavage-resistant. We suggest that dengue virus may have driven the long-term evolution of primate STING and driven more recent changes in chimpanzee STING.

Expanded summary*: My research has focused on new ways to identify immune factors that may be important for the pathogenicity of dengue virus in humans. I have implemented a number of different approaches including classic evolutionary methods of detecting recent and recurrent positive selection, as well as novel high-throughput machine learning techniques to identify new host genes that the dengue virus targets during infection. These approaches have been combined with experimental validation in the lab. It is my hope that this work will help expand the knowledge of this human pathogen, and how we may combat its endemic spread in humans.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-324

Genetic analysis of transmission stage production in schistosome parasites

Winka Le Clec'h [1,*], Frédéric D. Chevalier [1], Marina McDew-White [1], Vinay Menon [1], Tim J. C. Anderson

[1]Genetics, Texas Biomedical Research Institute, San Antonio, United States

Abstract: Parasite traits associated with transmission success, such as the number of infective stages released from the host, are expected to be optimized by natural selection. However, in the trematode parasite *Schistosoma mansoni*, a key transmission trait – the number of cercariae larvae shed from infected *Biomphalaria spp*. snails – varies significantly within and between different parasite populations and selection experiments demonstrate that this variation has a strong genetic basis. We used genetic crosses to determine the genetic architecture of this critical transmission related life-history trait. A *S. mansoni* isolate from Brazil (SmBRE) sheds very low numbers of cercariae, and causes minimal mortality to snails, while another new world parasite (SmLE) sheds 8 fold more cercariae (mean (±se) cercariae per shedding: 284±19 vs 2352±113) and causes high mortality to snails. We conducted two independent three generation genetic crosses between these two parasite lines (SmBRE and SmLE), and determined shedding profiles of parent parasites, F1 and F2 progeny from inbred *B. glabrata* snails. We sequenced the ~15 Mb exomes from parents, F1 progenitors and 188 F2 progeny for each crosses, revealing 9,140 and 9,465 SNPs fixed for alternative alleles in the two crosses, and conducted a classical QTL (*i.e.* Quantitative Trait Locus) analysis. The QTL analysis revealed potential QTLs on chromosome 1. We are now in position to identify candidate gene(s) involved in a key life-history trait that is critical for transmission in an important human pathogen.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

OTH-HP5

Using historical collections to understand the evolutionary response of bees to an emergent parasite

Katarzyna Bozek, Juliana Rangel, Jatin Arora, Mandy Tin, Emily Crotteau, Gerald Loper, Jennifer Fewell, Alexander Mikheyev*

Abstract: Populations must rapidly respond to sudden biotic pressures, such as those from novel diseases and pathogens, or face potential extinction. Despite continuous biotic challenges faced by most organisms, the genetic architecture of adaptation to parasites and diseases remains poorly understood, particularly in natural environments. Direct measurement of evolutionary changes caused by biotic pressures requires long-term field studies, which must also span a selection event. Here we take advantage of unique decade-long data sets from two wild honey bee populations in the U.S., to reconstruct evolution of tolerance to a novel parasite, the *Varroa* mite. After a host switch in the early 20th century, *Varroa* started a worldwide pandemic, which led to widespread honey bee colony collapses, though some populations are known to have evolved certain level of tolerance. The study populations simultaneously suffered massive *Varroa*-induced mortality in 1996, but stabilized within two years. Using sequenced and phased genomes of 465 specimens, we inferred the strength of selection acting on every polymorphic gene in naive and adapted populations. Despite massive gene flow from Africanized bees during the same time frame, selection acted primarily on standing genetic variation, with immigrant alleles playing only a minor role. Remarkably, gene-wise changes were strongly correlated across the separate populations, indicating parallel selective responses to *Varroa*. Genes involved in metabolism, protein processing and development were under particularly strong selection before and after the arrival of *Varroa*. Interactions among the highly connected genes in these pathways may channelize selective responses, causing unrelated populations to exhibit parallel evolutionary trajectories when faced with the same challenge. Our analyses illustrates that ecologically relevant traits result from highly polygenic selection involving thousands of genes in the genome.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-337

Truly Ubiquitous, CRESS DNA Viruses Scattered Across the Eukaryotic Tree of Life

Lele Zhao [1,*], Siobain Duffy [1]

[1]Department of Ecology, Evolution and Natural Resources, Rutgers University, New Brunswick, United States

Abstract: Until recently, viral detection and characterization were limited to viruses of economic significance, mostly associated with agriculture and disease pathogens. This bias caused virologists to think of the circular Rep-encoding single-stranded DNA (CRESS DNA) viruses as a relatively small group of viruses. However, with the explosion of metagenomic sequencing over the past decade, and with increasing use of the highly processive phi29 polymerase for rolling-circle amplification, scientists have identified and annotated copious amounts of novel CRESS DNA viruses in various ecosystems on all continents.
While harnessing host machinery for replication, DNA viruses have been known to integrate into host genomes. The Rep protein (replication-associated protein) is a conserved protein shared by the CRESS DNA viruses to assist replication. Detecting endogenous sequences homologous to viral Reps will not only provide us with "fossil records" for protein evolution studies but also reveal historical host species of these viruses. A systematic search for endogenous Rep sequences in the GenBank non-redundant eukaryotic protein sequence database was performed using Blastp. We utilized relaxed search criteria to capture divergent integrated Rep proteins. 314 unique hits resulted from the search, of which 146 were from the RefSeq protein database. The endogenous Rep fragments came from species of 26 different phyla across the eukaryotic tree of life. Maximum likelihood trees were constructed with bootstrap support depicting different species of the same viral family grouping with different endogenous Rep sequences.

Expanded summary*: Viruses are ubiquitous and can infect organisms all across the tree of life. Until recently, viral detection and characterization were limited to viruses of economic significance, mostly associated with agriculture and disease pathogens. This bias caused virologists to think of the circular Rep-encoding single-stranded DNA (CRESS DNA) viruses as a relatively small group of viruses. However, with the explosion of metagenomic sequencing and increasing use of the highly processive phi29 polymerase for rolling-circle amplification over the past decade, scientists have identified and annotated copious amounts of novel CRESS DNA viruses in various ecosystems on all continents, such as lakes, oceans, plant tissues, and mammal feces. The number of publications on novel CRESS DNA viruses has rocketed to almost three digits over the past seven years with no sign of plateauing.
While harnessing host machinery for replication, DNA viruses have been known to integrate into host genomes. The Rep protein (replication-associated protein) is a conserved protein shared by all CRESS DNA viruses to assist replication. A systematic search for endogenous Rep sequences in the GenBank non-redundant eukaryotic protein sequence database was performed using Blastp. We utilized relaxed screening criteria to capture divergent integrated Rep protein within eukaryotic genomes. The BLOSUM50 matrix was used with an e-value threshold of 0.001, gap penalty of 15 and extension penalty of 1. Three hundred and fourteen unique hits resulted from the search of the GenBank non-redundant eukaryotic protein database, of which 146 were from the RefSeq protein database. The shortest endogenous sequence fragment was 48 amino acids and the longest 440 amino acids. The endogenous Rep fragments came from species of 26 different phyla across the eukaryotic tree of life. Three maximum likelihood trees were constructed with bootstrap support depicting different species of Geminiviridae, Circoviridae and Nanoviridae grouping with different endogenous Rep sequences.
This survey of endogenous viral Reps suggested an unexpectedly wide distribution of endogenous Rep protein sequences across the eukaryotic tree of life. Mounting metagenomic viromes do not identify which viruses infect which hosts, although the species found containing endogenous Rep sequences qualitatively agree with the environments where novel CRESS DNA viruses are being discovered. The origin of these endogenous fragments could be used to narrow down candidate hosts for virus and host interaction studies of CRESS DNA viruses. Exposure of this link between viruses and eukaryotes created by an integrated conserved protein will provide insights into the evolutionary history of the Rep protein and CRESS DNA viruses.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-318

Divergence of HPV16 variants reflects loci undergoing inter-host positive selection, potentially immunologic selection

Chase Nelson [1,*], Apurva Narechania [1], Robert Burk [2], Mark Schiffman [3], Michael Cullen [3], Joseph Boland [3], Zigui Chen [4], Nicolas Wentzensen [3], Qi Yang [3], Jason Mitchell [3], David Roberson [3], Sara Bass [3], Laurie Burdett [3], Tina Raine-Bennett [5], Thomas Lorey [6], Philip Castle [2], Meredith Yeager [3], Lisa Mirabello [3]

[1]Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, [2]Albert Einstein College of Medicine, Bronx, NY, [3]National Cancer Institute, Rockville, MD, United States, [4]The Chinese University of Hong Kong, Hong Kong, Hong Kong, [5]Kaiser Permanente, Oakland, CA, [6]Kaiser Permanente, Berkeley, CA, United States

Abstract: Papillomaviruses are one of the most successful families of vertebrate DNA viruses. Among them, human papillomavirus type 16 (HPV16) is the most carcinogenic, causing approximately 50% of all cervical cancers. Unfortunately, no straightforward phylogenetic relationship or genetic variant(s) explains HPV oncogenicity, e.g., the second most carcinogenic type (HPV18, causing ~16% of cancers) is relatively distantly related to HPV16. Thus, the genetic and evolutionary mechanisms underlying HPV16's unique carcinogenicity remain unsolved. Mirabello et al. recently reported on viral genome data from 3,215 HPV positive specimens from women undergoing cervical cancer screening at Kaiser Permanente Northern California in the Persistence and Progression (PaP) cohort. Results demonstrated profound differences in disease risk by histologic subtype among the HPV16 sublineages (e.g., A1, A2, and D2), with over 100-fold risk differences. Here, we expanded this analysis to examine the genetic and evolutionary underpinnings of HPV16 carcinogenicity through an exhaustive analysis of viral nucleotide diversity.

The HPV16 genome displays significant evidence of purifying selection (d_N/d_S = 0.267; P<0.001). However, within the A1 and A2 sublineages, one of the two oncogenes (E6) does not differ from neutrality, while the other oncogene (E7) is significantly more constrained in cases (d_N/d_S = 0.049) than in controls (d_N/d_S = 0.27; P<0.001). Thus, benign viral infections exhibit less constraint, implying a nonsynonymous mutational burden. Using an unsupervised sliding window approach, we next identified genomic regions exhibiting strong evidence of inter-host positive selection. Among all of the sublineages, 26 regions displayed d_N/d_S values ranging from 1.28 – 33.52. A subset of these regions overlapped with phylogenetic sublineage-defining residues, therefore, we further analyzed individual sublineages to control for potential lineage fixation by genetic drift. Remarkably, 13 of these 26 regions were discovered independently in the A1 sublineage alone, with d_N/d_S values up to 59.44. Nine of these 13 regions match known HPV epitope sequences obtained from the Immune Epitope Database and/or the Human Papillomavirus T Cell Antigen Database. In particular, two regions of E6 were identified independently in cases and controls: codons 20-27 and codons 75-90, which include four sublineage-defining residues. Both overlap with experimentally verified HLA class I epitopes, and the second region also overlaps known antibody epitopes. Moreover, the 75-90 region exhibits a d_N/d_S ratio of 59.44 in cases and 28.05 in controls, shows substantial case vs. control divergence (between – within group d_N/d_S of 55.60), and includes a L83V variant that has been reported to contribute to persistence in European populations.

We conclude that balancing positive selection likely played a key role in the historical divergence of HPV16 variants, possibly as the result of inter-host environmental heterogeneity based on host immune genotype. Moreover, since our data imply that positive selection is targeted to many of the same loci that are diagnostic as sublineage-defining residues, it is likely that similar evolutionary pressures have operated throughout the evolutionary diversification of HPV16. In particular, we suggest that host immune genotype (e.g., HLA) may play a key role in disease outcome, and must be prioritized in future studies of HPV evolution and its link to cervical cancer.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-328

Genetic Variation in Host Tolerance of an Invading Transposable Element

Erin Kelleher [1,*], Jyoti Lama, Uche Akoma [1], Lily Ortega [1]

[1]Biology and Biochemistry, University of Houston, Houston, United States

Abstract: Transposable elements (TEs) are genetic parasites that harm their host by inducing deleterious mutations and genomic instability. To avoid these fitness costs, host repression of germline TE activity is known to evolve through small RNA mediated silencing pathways in both plants and animals. However, it remains unknown whether host genomes may also evolve tolerance of TEs, by desensitizing themselves to the harmful consequences of TE activity. To this end, we performed a genome wide association study of tolerance of *P*-element DNA transposons in *Drosophila melanogaster*. Because *P*-elements are regulated by piRNAs that are transmitted exclusively through the female germline, we were able to examine tolerance of *P*-elements in the dysgenic offspring that are produced when males bearing active *P*-elements are mated to naive females. By crossing panel of genotyped naïve recombinant inbred lines to males from a reference strain that exhibits strong P activity, we uncovered multiple QTL that determine tolerance. Candidate loci under these QTL imply that tolerance of P activity is conferred through variants in protein coding genes that minimize the disruption of gametogenesis, and through the dosage or regulation of other selfish genetic elements that may be dysregulated in the dysgenic germline.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-320

Polymorphism in P-element repressor alleles

Shuo Zhang [1,*], Erin S. Kelleher [1]

[1]Department of Biology and Biochemistry, University of Houston, Houston, United States

Abstract: Transposable elements (TEs) are selfish genetic elements, whose mobilization can reduce host fitness by producing deleterious mutations and inciting genome instability. In many metazoans, TE activity is regulated in the germline by small Piwi-interacting RNAs (piRNAs), which are derived from specialized genomic loci known as piRNA clusters. When a new TE invades the genome via horizontal transfer, repression is proposed to evolve through *de novo* mutation, with random transpositions into piRNA clusters producing repressor alleles. Alternatively, repression may also evolve through epigenetic mutation, if an insertion of the invading TE is converted into a piRNA cluster.

P-elements are DNA transposons that invaded the *D. melanogaster* genome around 1950. In response, many natural populations of *D. melanogaster* evolved *P*-element repression extremely rapidly, in less than 50 years. Although piRNAs are proposed to enact *P*-element repression, the challenge of annotating TE insertions in heterochromatin, where piRNA clusters reside, has prohibited the identification of *P*-element repressor alleles (*i.e.* piRNA-producing insertions), which underlie the repressive phenotype. Using a novel approach, we successfully curated *P*-element insertions in at least one piRNA cluster in ~99% of 200 wild-derived genomes which comprise the *Drosophila* Genetic Reference Panel (DGRP). Our analyses uncovered no fewer than 10 distinct repressor alleles, indicating that the *de novo* mutation rate to repression is very high. Taken together, our results support transposition into existing piRNA clusters as the predominant mutational mechanism for the evolution of *P*-element repression, and reveal unexpected genetic complexity behind this rapid evolutionary transition.

Expanded summary*: Transposable elements (TEs) are selfish DNA segments that move and replicate within their hosts' genomes. They are extraordinarily evolutionarily successful, being present in almost all genomes, and comprising up to 80% of their hosts' genomes. However, TEs are also genomic parasites that harm host fitness by producing deleterious insertions, inducing DNA double-strand breaks and mediating ectopic recombination. These mutational effects are associated with severe phenotypic consequences such as sterility and cancer. Therefore, the host needs to control TEs, particularly in the germline, where TE activity impacts reproductive fitness, and deleterious insertions will be transmitted to the next generation.

Small RNA-mediated silencing is a conserved strategy for germline TE regulation. Small interfering RNAs (siRNAs) in plants and Piwi-interacting RNAs (piRNAs) in most animals target the transcriptional and/or post-transcriptional silencing of resident TEs, thereby restricting their mobilization and mutational effects. However, genomes are recurrently invaded by new TE families which are horizontally transferred, often from distantly related lineages. Two mutational mechanisms are proposed to introduce novel small RNAs that target an invading TE for silencing into the small RNA pool. First, small RNA-mediated silencing could be established by *de novo* mutation if an invading TE randomly transposes into an existing small RNA-encoding locus. Alternatively, silencing could be established by epigenetic mutation (epimutation) if non-small RNA producing insertions of an invading TE were converted into small RNA producing loci. Distinguishing between these models is challenging, because the identity of small RNA producing regions prior to a TE invasion are rarely known.

To unveil the contribution of *de novo* and epigenetic mutations to the evolution of small RNA-mediated silencing, I am taking advantage of the recent invasion of the *D. melanogaster* genome by *P*-element DNA transposons. *P*-elements invaded the *D. melanogaster* genome around 1950 and quickly spread worldwide. In many geographic locations, such as North America, repression evolved almost concurrently with invasion, and by ~1980 most strains exhibited strong repression. Repression is thought to be mediated predominantly by piRNAs, which are encoded by specialized heterochromatic loci called piRNA clusters. Because the *P*-element invasion happened on a historical time scale, ancestral piRNA clusters that pre-date the invasion are known, based on their presence in genomes that were collected prior to 1950.

To uncover the piRNA producing *P*-elements in natural populations, I took advantage of the *Drosophila* Genetic Reference Panel (DGRP), a group of 205 fully sequenced inbred lines, which were collected from the Raleigh, North Carolina, USA in 2003. Previous TE annotations suggest that fewer than 25% of DGRP genomes harbor *P*-element insertions in ancestral piRNA clusters, which would imply a major role for epimutation in the evolution of *P*-element repression. I present a new method for annotating TEs in repetitive regions and satellite blocks, which relies on their unique assignment to a particular genomic region rather than an exact nucleotide position. Using this new approach, I was able to identify at least one *P*-element insertion in an ancestral piRNA cluster in ~99% of the DGRP genomes. Furthermore, I have identified at least 10 distinct repressor alleles (independent insertion of *P*-elements into piRNA clusters) among the DGRP panel. My results suggest that transposition of invading TEs into existing small RNA-producing regions is the predominant mechanism for TE regulation, and that the *de novo* mutation rate to piRNA-mediated repression is very high. By comparing allele frequencies of *P*-element insertions within and outside of piRNA clusters, I am further testing if positive selection was involved in the evolution of piRNA-mediated *P*-element regulation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-327

Rapid evolution of the mammalian heat-stable enterotoxin receptor and its peptide hormone ligand

Clayton Carey [1,*], Nels Elde [1]

[1]Human Genetics, University of Utah, Salt Lake CIty, United States

Abstract: Pathogen-encoded mimicry presents an evolutionary obstacle for hosts to avoid pathogenic exploitation while executing core functions. Characterization of genetic conflicts at interfaces involving pathogen mimicry of host factors revealed notable cases of constraint for retaining host affinity to ligands despite adaptation to rapidly evolving mimics. These findings suggest limited capacity for co-evolution at host-host interfaces exploited by pathogen mimicry. Here we describe a core physiological function marked by rapid evolution of both host receptor and its ligand, despite pathogenic mimicry of the ligand. The signaling protein Guanylate Cyclase-C (GC-C) promotes water secretion in the intestinal epithelia through binding of the peptide hormone uroguanylin. Some bacterial pathogens hijack GC-C signaling by producing heat-stable enterotoxins (ST), which mimic features of uroguanylin to bind GC-C and promote spread through watery diarrhea. Sequence comparisons revealed signatures of positive selection in GC-C among primates and among bats. In both lineages, positively selected sites are limited to the toxin interacting extracellular domain of GC-C adjacent to critical ligand binding residues. Experimental characterization of primate GC-C revealed species-specific differences in susceptibility to ST variants encoded by clinically-derived *E. coli* strains, consistent with a history of pathogen-driven evolution. Characterization of GC-C from the vesper bat *Myotis lucifugus* also revealed significant differences in ST susceptibility compared to human GC-C. Surprisingly, while human GC-C is activated by uroguanylin from distantly related mammals, *M. lucifugus* GC-C was not activated by human uroguanylin, suggesting that GC-C and uroguanylin may have undergone lineage-specific co-evolution in vesper bats. Together, these data reveal that diarrheal pathogens may influence the evolution of both a receptor and ligand from a conserved host interface, as part of an ongoing evolutionary battle for control of water physiology in the gut.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POA-419

INTERACTIONS BETWEEN SMALL RNA AND TRANSPOSABLE ELEMENT EXPRESSION IN DROSOPHILA

Andrea Betancourt*, Olga Pawlowska [1]

[1]University of Liverpool, Liverpool, United Kingdom

Poster:

Transposable elements are remarkably successful genomic parasites that form a substantial fraction of most eukaryotic genomes. They spread by occasional horizontal transmission, followed by propagation within genomes via new germline insertions. To protect the integrity of their genomes, hosts have evolved a number of defense mechanisms, including epigenetic suppression of transposable element expression and degradation of transposable element transcripts. Here, we examine interactions between the expression of transposable elements and the regulatory RNAs targeting them. We find that production of more transposable element message leads to greater numbers of small RNAs that target them. We also explore the transcriptome for factors that protect against germline dysgenesis, which is due to the expression of a particularly destructive transposable element, the P-element.
Transposable elements are remarkably successful genomic parasites that form a substantial fraction of most eukaryotic genomes. They spread by occasional horizontal transmission, followed by propagation within genomes via new germline insertions. To protect the integrity of their genomes, hosts have evolved a number of defense mechanisms, including epigenetic suppression of transposable element expression and degradation of transposable element transcripts. Here, we examine interactions between the expression of transposable elements and the regulatory RNAs targeting them. We find that production of more transposable element message leads to greater numbers of small RNAs that target them. We also explore the transcriptome for factors that protect against germline dysgenesis, which is due to the expression of a particularly destructive transposable element, the P-element.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-332

Cynomorium coccineum, a parasite species that acquired genes from different hosts as it spread from Mongolia to the Mediterranean

Susanne Renner*, Natalie Cusimano [1]

[1]Systematic Botany and Mycology, University of Munich (LMU), Munich, Germany

Abstract: *Cynomorium coccineum* is a holoparasite that occurs from the Mongolian deserts in western China through Iran and Saudi Arabia to the Canary Islands. It belongs to the Saxifragales and parasitizes species from several families, including Amaranthaceae, Chenopodiaceae, Tamaricaceae (all Caryophyllales), and Nitrariaceae (Sapindales), some of them only found in parts of its range. This offers an opportunity to detect multiple horizontal gene acquisitions by a single species as it expanded its geographic and host range. We analyzed numerous genes from the *Cynomorium* nuclear, plastid, and mitochondrial genomes, and their homologues in host individuals from China, Algeria, Morocco, Sicily, Spain and Portugal and thereby were able to unravel sequential HGT over space and time.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-329

Iron and zinc piracy: More evidence of co-evolution of bacteria and primates

Klim Kostyuk [1], Georgii Bazykin [2 3], Sofya Garushyants [3,*]

[1]Moscow Gymnasium in the South-West No.1543, [2]Skolkovo Institute of Science and Technology, [3]IITP RAS, Moscow, Russian Federation

Abstract: Transitional metals are vital for all life forms. Pathogenic bacteria have to harvest these metals from their hosts. To restrict free metals concentration in body fluids mammals developed mechanisms of nutritional immunity. Most of iron ions are bound to transferrin, lactoferrin, and hemoglobin, and zinc ions – to calprotectin. Bacteria by the same token developed mechanisms to overcome this by producing various proteins capable to extract bound metal ions and transfer them to bacterial cells. It was shown previously, that genetic variation in transferrin among primates is the result of rapid evolution aimed to resist bacterial metal piracy by transferrin binding protein A (TbpA).

In the present work other systems involved in nutritional immunity, such as transferrin and TbpB; lactoferrin and LbpA; hemoglobin and HpuA and HpuB; calprotectin and CbpA; and FrpB were investigated. To find sites under positive selection we utilized either a dataset of 16 primates from UCSC, or a set of representative strains for four bacterial species (*Neisseria meningitidis*, *Neisseria gonorrhoeae*, *Haemophilus influenzae* and *Haemophilus parasuis*). We found rapidly evolving residues in transferrin that interacted with residues under positive selection in TbpB, and two sites under positive selection in hemoglobin that interacted with HpuA, that provided additional evidence for Red Queen hypothesis.

Expanded summary*: My research is focused on different aspects of bacterial evolution: interspecies interactions in microbial communities, and bacterial adaptation to the host. Host-bacterial interactions are of the particular significance, because understanding of such interactions in pathogenic bacteria may provide clues for better treatment of bacterial diseases and gives crucial knowledge for understanding the mechanisms of bacterial speciation.

In the work submitted to SMBE the competition for nutritional metals between pathogenic bacteria and primates was investigated, and we were able to find rapidly evolving interacting sites in both primates, and bacteria in complexes of hemoglobin and HpuA, and transferrin and TbpB. These findings provide more evidence for the Red Queen hypothesis.

There are other curious cases of adaptation to the host in bacteria, e.g. changes in bacterial metabolism and protein composition in response to rich and stable environment inside the host. In other part of my research I studied metabolism of *Holospora* spp., endonuclear symbionts of Paramecium (joint project with Department of Genetics and Selection, SPSU). We have shown, that in order to survive in the nucleus these bacteria developed unique metabolism. They have lost most central metabolic pathways, but instead they get all the energy from ribonucleotides, and steal aminoacids from the host. Another case that I have investigated concern the changes in ribosomal composition in bacteria with reduced genomes. We found some distinctive patterns of ribosome degradation among a set of various bacteria.

One more interesting aspect of rapid evolution is horizontal gene transfer (HGT). At first we estimated a portion of genes in methanogenic Archaea horizontally transferred from bacteria. We confirmed the previous observations that HGT from bacteria played a crucial role in the origin of methanogenic Archaea, and provided more conservative estimate of the amount of transferred genes. What is more, we tried to estimate the size of horizontally transferred regions and found out that most of them were rather short. HGT may as well play a crucial role in pathogenesis. In the joint project with SRI PCM we investigated Crohn's disease associated *E.coli*, and found that some of the strains isolated from the patients with Crohn's disease carry plasmids highly similar to that of pathogenic bacteria. We consider that acquisition of these plasmids may lead to pathogenicity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-330

Population dynamics of antibody repertoires in response to HIV-1 infection

Armita Nourmohammad [1,*], Jakub Otwinowski [2], Marta Luksza [3], Thierry Mora [4], Aleksandra Walczak [4]

[1]Princeton University, Princeton, [2]University of Pennsylvania, Philadelphia, [3]institute for advanced study, princeton, United States, [4]École Normale Supérieure, Paris, France

Abstract: We normally think of evolution occurring in a population of organisms, in response to their external environment. Rapid evolution of cellular populations also occurs within our bodies, as the adaptive immune system works to eliminate infection. Chronic pathogens, such as HIV, are able to persist in a host for extended periods of time, during which they also evolve to evade the immune response. In HIV infected individuals, affinity maturation and hypermutation of B-cell receptors creates large amounts of genetic diversity and turnover. We study the dynamics of B-cell repertoires in a number of untreated HIV infected individuals from sequences collected over 2 years. Using a probabilistic scheme, we infer the clonal B-cell lineages within a repertoire and construct maximum likelihood phylogenies for each lineage. Based on statistics of mutations along the phylogeny, we show evidence for strong positive selection on non-synonymous mutations in CDR regions, coupled with negative selection on the framework region of B-cell receptors in HIV patients. We infer competition between B-cells within lineages, and demonstrate that clonal competition and turnover of B-cell mutants during maturation of a lineage are predictive of HIV-specific responses in the repertoire. We validate our results by comparison with unproductive repertoires, drug treated HIV infected individuals, and with healthy individuals.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-340

Great ape genetic diversity suggests different evolutionary trajectories of innate immunity genes

João C. Teixeira [1 2,*], Athanasios Kousathanas [1 2], Étienne Patin [1 2], Guillaume Laval [1 2], Lluis Quintana-Murci [1 2]

[1]Unit of Human Evolutionary Genetics, Institut Pasteur, [2]CNRS URA, Paris, France

Abstract: Innate immunity constitutes the front line of host defence against pathogens, and increasing evidence suggests that humans and other apes exhibit important differences in susceptibility to infectious diseases (e.g. HIV, malaria). Nevertheless, the role played by selection in species-specific adaptations to pathogen pressure remains poorly understood. In this study, we uncovered shared and species-specific signatures of natural selection acting on innate immunity genes in great apes at different time scales, from thousands to millions of years of evolution. We analyzed whole-genome sequence on a total of 145 samples from 9 great ape species, and compared selection signatures on a set of ~1,500 innate immunity genes to the remainder of the genome. We focused on three forms of natural selection: purifying selection by estimating the the distribution of fitness effects of new non-synonymous mutations in each species; balancing selection by measuring the excess of genetic diversity and the dispersion of allele frequencies from expectations under a demographic model with purifying selection; and positive selection by combining a set of summary statistics based on haplotype homozygosity and population differentiation. Our analyses unveil important differences in the mechanisms involved in host adaptation to pathogen pressures across species: while targets of both purifying and balancing selection are more commonly observed in more than one species, genes affected by positive selection are often species-specific. To our knowledge, this study represents the first attempt to reconstruct the evolutionary mechanisms that operated for millions of years of primate evolution as a response to pathogen infection.

Expanded summary*: Innate immunity constitutes the front line of host defence against pathogens, and provides a valuable model for the study of the selective pressures imposed by microorganisms on host genomes. Population genetic studies in humans have shown that the impact of selection on some families of innate immune receptors and downstream signaling molecules varies considerably. Although humans and closest relatives share most of their genome, increasing evidence suggests that humans and other apes exhibit important differences in susceptibility to, and severity of, infectious diseases (e.g. HIV, malaria). These phenotypic differences between species could arise by several types of selection acting on genetic variation of innate immune genes: relaxed purifying selection in some lineages due to low effective population size, positive selection leading to species-specific adaptations in response to pathogen challenges, and balancing selection preserving genetic diversity that could allow populations that experience temporally and spatially varying pathogenic pressure to survive. However, the extent to which different types of selection have shaped innate immunity in our closest relatives remains largely unknown.

In this study, we aimed at uncovering shared and species-specific signatures of natural selection acting on innate immunity genes in great apes. We implemented a strategy that allowed us not only to investigate different forms of natural selection, but also to unveil examples where selection has acted at vastly distinct evolutionary times, varying from thousands to millions of years. We analyzed whole-genome sequence data for different great ape populations, covering a total of 145 samples from 9 species, and focused on comparing selection signatures on a set of 1,553 genes, involved in innate immunity functions, to the remainder of the genome. We started by estimating, for the first time, the distribution of fitness effects of new non-synonymous mutations (DFE) in each great ape species, accounting for their respective demographic history, which allowed us to evaluate the strength of purifying selection acting on all genes and compare to different classes of innate immunity genes. We then used this information to investigate the role of balancing selection in these genes, by measuring the excess of genetic diversity and the dispersion of allele frequencies from expectations under a demographic model with purifying selection. Furthermore, we uncovered targets of positive selection among innate immunity genes by focusing on the most extreme signatures genome-wide, as measured by a Fisher's score combining a variety of summary statistics, based on haplotype homozygosity (iHS, DiHH and XP-EHH) and on population differentiation (FST and DDAF). Our analyses show that some loci exhibit extensive differences in the form of selection across species: while targets of both purifying and balancing selection are more commonly observed in more than one species, genes affected by positive selection are often species-specific. Among shared targets of balancing selection across species, we found CD36, an immune receptor associated with the phagocytic uptake of malaria (among other pathogens) that has been proposed to evolve under positive selection in human populations from Africa.

These findings suggest important differences across primates in the mechanisms involved in host adaptation to pathogen pressures, informing about the pressures imposed by their respective ecological habitat. To our knowledge, this study represents the first attempt to reconstruct the evolutionary mechanisms that operated for millions of years as a response to pathogen infection.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-338

Comparative genomics of symbiont Cardinium species reveal the asexuality of host mites

Phuong Le [1 2,*] and Zaichao Zhang, Yao-Cheng Lin, Juliana De Freitas Astua, Denise Navia Magalhaes Ferreira, Johannes AJ Breeuwer, Thomas Van Leeuwen, Yves Van de Peer

[1]Center for Plant Systems Biology, VIB, Ghent University, Technologiepark 927, [2]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Gent, Belgium

Abstract: The false spider mite Brevipalpus's parthogenesis was induced by bacterial symbionts. The hypothesis is that the obligate intracellular bacterium - Cardinium species - induced the asexual reproduction of the host mites. Cardinium infected haploid female reproduces only female progeny, and non infected female reproduces few male progeny. Moreover, the phylogenetic evidences showed that these bacteria must have been laterally transferred between mite clonal lineages and may facilitate the laterally gene transfer between mite hosts. Studying Brevipalpus genomes along with their endosymbiont bacteria can highlight which evolutionary factors make changes in the genetic structures and the differences in the asexuality of different mite clonal lineages.

The genomes of Brevivalpus were sequencing and annotating: B. yothersi, B. papayensis and B. obovatus. The bacterial sequences separated from their host genomes showed that each mite was contaminated by a different strain of Cardinium species. Cardinium of B.yothersi and B.obovatus belong to the same clade whereas Cardinium of B. payayensis belong to another clade based on the different phylogenetic trees including gyrA, gyrB and single core genes. This observation reveals that host mite could harbor either multi-infections or the integration of Cardinium genome into the host. Moreover, Cardinium infected Brevipalpus caused parthenogenesis in host while Cardinium infected Bermisia species and Encarsia species caused cytoplasmic incompatibility in host. The later Cardinum genomes were published two years ago. The comparative genomics of the different Cardinium genomes are under-investigating in order to clarify our hypothesis.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

OTH-HP4

Network Architecture is a Fundamental Constraint on Optimal Immunity during Coevolution with a Signal-Disrupting Parasite

Edward Schrom [1,*], Joaquin Prada [2], Andrea Graham [1]

[1]Ecology and Evolutionary Biology, Princeton University, Princeton, United States, [2]Mathematics Institute, University of Warwick, Warwick, United Kingdom

Abstract:

Defense against infection incurs costs and benefits that shape the evolution of defense strategies, and many theoretical studies have addressed contexts in which hosts are expected to adopt constitutive versus inducible defenses. These strategies differ in the timing of immune readiness: immediate (for constitutive) or delayed, even if briefly (for inducible). However, even when one immune timing strategy is theoretically optimal, it may be evolutionarily unachievable. This is because evolution proceeds via stepwise mutational changes to the protein interaction networks underlying immune responses, not by changes to immune timing directly. Here we use a theoretical simulation model to examine how underlying network architectures constrain the evolution of immune timing strategies, and how these architectures confer desirable immune properties such as inducibility and robustness. We focus on immune signaling because signaling molecules are common targets of immune subversion but are rarely studied in this context. We find that in the presence of a coevolving parasite that disrupts immune signaling, hosts evolve constitutive defenses even when inducible defenses are theoretically preferable. This is because there are relatively few network architectures capable of producing inducible immunity that are robust to targeted disruption, and because evolution towards robust inducible network architectures often requires intermediate steps that are vulnerable to targeted disruption.

Expanded summary*:

Optimal immune systems are those which best balance numerous and conflicting selective pressures [1]. Besides clearing infections rapidly, these selective pressures include minimizing metabolic costs [2], minimizing the risk of immunopathology [3], and maintaining functionality despite parasite sabotage [4]. Combating sabotage, which we explicitly examine, is a particularly important but understudied aspect of host-parasite coevolution [5].

Accounting for these selection pressures, many theoretical studies have addressed the ecological scenarios that favor some immune strategies over others [6,7,8,9,10,11]. Frequently investigated are immune timing strategies – that is, constitutive (constantly produced and therefore immediately effective upon infection) vs. inducible defense (produced only upon detecting an infection and therefore delayed). Theoretical studies typically represent the constitutive and inducible components of immune defense as state variables or parameters, implicitly assuming that the optimal strategy will be achieved. However, an immune strategy may be optimal in theory but unattainable in practice. To study the evolvability of immune timing strategies, we must recognize that timing is an emergent observation produced by individual molecules and their network of interactions. It is this molecular playing field upon which genetic mutations have their immediate effects, and upon which selection operates.

To study how the architectures of molecular networks underpin the evolution of immune timing, we simulate the evolution of immune protein networks. We assume that rapidly inducible defense is theoretically optimal [8], but such timing must result from the underlying protein network. Because immune signaling proteins are understudied in the context of parasite sabotage and yet seem to be most often targeted by parasite disruption strategies [12], we compare purely evolved immune networks to networks coevolved with signal-disrupting parasites. The coevolution scenario selects not only for inducible immune networks, but also for networks that are robust against targeted disruption.

We find that in the pure evolution scenario, host populations achieve optimal inducible immunity. To the contrary, in the coevolution scenario, host populations rely on constitutive defense, even though inducible immunity is theoretically optimal. This occurs for two reasons, both of which involve the underlying signaling networks. First, the set of network architectures producing robust inducible immunity is smaller than the set producing robust constitutive immunity. Mutations occur randomly, so hosts will approach and attain robust constitutive immunity more frequently than robust inducible immunity. Because coevolution selects for

robustness, host populations are unlikely to leave a state in which all hosts have achieved robust immune defense, even if this immune defense is suboptimal. Second, even when mutations allow hosts to approach inducible immunity, this often proceeds through intermediate networks that are vulnerable to targeted disruption. Thus, each approach to robust inducible immunity is easily blocked by coevolving parasites, and the host population settles for robust constitutive immunity instead. Throughout this work, we also identify specific structural features of protein networks that are responsible for the emergent inducibility and robustness of immunity.

This work highlights the importance of the protein interaction networks responsible for the emergent immune properties that are of great interest to evolutionary biologists. Thus, our modeling framework links the often disparate fields of molecular immunology and evolution. As molecular biologists decipher the network architectures of model organism immune systems [13,14], this framework can help predict the coevolutionary ramifications of mutations to different components of the network. Moreover, as evolutionary biologists characterize the coevolutionary battles between hosts and parasites, this framework can help explain the fundamental constraints on optimal immunity imposed by its molecular underpinnings.

References: [1] Sheldon & Verhulst. (1996) Trends Ecol Evol. [2] Lochmiller & Deerenberg. (2000) Oikos. [3] Graham et al. (2005) Annu. Rev. Ecol. Evol. Syst. [4] Bergstrom & Antia. (2006) Trends Ecol Evol. [5] Schmid-Hempel. (2008). Trends Ecol Evol. [6] Shudo & Iwasa. (2001) J. Theor. Biol. [7] Shudo & Iwasa. (2002) J. Theor. Biol. [8] Frank. (2002) J. Theor. Biol. [9] Hamilton et al. (2008) Proc. R. Soc. B. [10] Ito & Sakai. (2009). J. Theor. Biol. [11] Kamiya et al. (2016) BMC Evol. Biol. [12] Lazzaro. (2008) Curr. Opin. Microbiol. [13] Teixeira. (2012) Brief. Funct. Genomics. [14] Kim et al. (2014) Cell Host Microbe.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

OTH-HP1

Empirical examination of the genomic signatures of host-parasite coevolution in the Diversity Panel of Daphnia magna

Peter Fields [1,*], Dieter Ebert

[1]Zoology Institute, University of Basel, Basel, Switzerland

Abstract: Hosts evolve to minimize the fitness reduction caused by parasites, while parasites optimize the exploitation of their hosts. In models of this process high genetic specificity in host–parasite interactions is assumed. These interactions are in the center of the theory of host–parasite coevolution and determine important aspects of the coevolutionary process, such as its tempo and mode, the occurrence of cyclic allele frequencies, and the potential for evolutionary novelty. These expectations can be condensed into distinct models of coevolutionary dynamics, e.g. arms-race and Red Queen/trench warfare models, which in idealized circumstances will provide succinct predictions for how pertinent genomic polymorphism will be distributed across a species' range. We provide a comprehensive description of phenotypic and genetic diversity, both at genome wide and putative loci associated with pathogen resistance, in the crustacean Daphnia magna, interacting with the bacterial pathogen Pasteuria ramosa and the microsporidians Ordospora colligata and Hamiltosporidium tvaerminnensis. Our high-throughput sequencing data from the Holarctic D. magna Diversity Panel and a well characterized Finnish rock-pool metapopulation, combined with long-read PacBio sequencing and de novo assembly of multiple clonal genotypes show the distinct role coevolutionary processes have on structuring genomic polymorphism while simultaneously revealing the limitations of individual approaches for inferring the (co-)evolutionary consequences of host-parasite interactions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-333

CCR5 is under recurrent positive selection, making some primates resistant to infection by immunodeficiency viruses (HIV and SIV)

Maria E Kaczmarek [1,*], John F Nahabedian [2], Emily R Feldman [3], Cody J Warren [3], Nicholas R Meyerson [3], Julie Overbaugh [2], Sara Sawyer [3]

[1]Integrative Biology, University of Texas at Austin, Austin, [2]Human Biology Division, University of Washington - FHCRC, Seattle, [3]MCDB, University of Colorado Boulder - BioFrontiers Institute, Boulder, United States

Abstract: CD4 and CCR5 are the receptor and co-receptor that allow human and simian immunodeficiency viruses (HIV and SIV) to enter cells. CD4 has been shown to be under recurrent positive selection, which has contributed to differences in primate susceptibility to immunodeficiency virus entry. The evolution and functional relevance of species-specific differences in CCR5 has largely been overlooked, partly due to the primary focus on CD4 as the cellular receptor responsible for differences in viral entry. Here, we performed an evolutionary analysis with a large dataset of primate CCR5 orthologs and show that primate CCR5 also bears the signature of recurrent positive selection, particularly within the New World Monkey clade. In addition, we have run site-specific models of positive selection that identified sites in the N-terminus of CCR5, which are highly variable within New World Monkeys. Amino acids in this region are sulfated and this modification is known to be required for viral entry. Interestingly, we show that the sulfation motif is absent in New World monkeys, and this contributes to lower levels of entry when a permissive CD4 is paired with a CCR5 from New World monkeys. New World Monkeys are not known to harbor any modern viruses related to HIV and SIV, this suggests the compelling hypothesis that New World monkeys have acquired adaptive mutations in CCR5 that protect them from these viruses. Together, our work shows evidence that CCR5 can pose as a barrier to cross-species transmission in primates.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

OTH-HP3

Extensive Recombination During Colonization of Cervid Endogenous Gammaretrovirus and Implications to Host Genome Evolution

Lei Yang [1,*], Mary Poss [1]

[1]Department of Biology and Center for Infectious Disease Dynamics, The Pennsylvania State University, University Park, United States

Abstract: Endogenous retroviruses (ERVs) are derived from infectious retroviruses that integrated into their host germline. Once integrated, ERVs are generally quiescent, but can occasionally impact the structure and function of the host genome. It is unclear whether newly acquired CrERVs could introduce diversity to the host genome, modulating its evolution. While most vertebrate hosts acquired ERVs in the ancient past, mule deer genomes are currently being colonized by the recently reported cervid endogenous gammaretrovirus (CrERV), which is still transcriptionally active and insertionally polymorphic in the population. We ask whether the evolutionary dynamics of newly acquired CrERVs implicate a role in the host genome evolution during the colonization period. We developed a draft mule deer genome and determined the location and sequence of all CrERVs in a single animal. Multiple waves of infection from 700,000 years ago to the present brought four distinct lineages of CrERVs in the contemporary host genome. Only the youngest lineage contains extant CrERVs without evidence of recombination, whereas recombination is prevalent within and between all CrERV lineages and sometimes involves host genome sequences. Recombinant activity at integration sites is also evident: ~37% of CrERVs lost their genes via recombination between the two homologous ends. CrERVs can potentially regulate host genes: ~30% of CrERVs integrated within 10kb of a host gene, ~30% of which are intronic. Overall, our data show that interaction within and between recently colonized ERV lineages is evident and suggest that the ERV colonization period provides an opportunity to alter the host genome evolution.

Expanded summary*: As horizontally acquired new genetic material, ERVs are known to have significant impact on the host genome: they can be co-opted to provide function to the host such as providing long non-coding RNA and protein-coding genes, they can regulate host genes by offering regulatory elements and modifying epigenetic configuration, they can also alter the host genomic structure by serving as recombination substrates. Because of these impacts, ERVs are under tight host regulation, and it is deemed that the majority of newly inserted ERVs are silenced by host defense mechanisms such as epigenetic regulation and antiviral factors. Evolutionary dynamics of newly acquired ERVs was formerly not testable because the majority of ERV colonization events are ancient. Yet ERVs could potentially affect host genome evolution at multiple times of a species evolutionary history: an ERV could be beneficial at one point but not at later times, and a colonized ERV could be neutral at first and then acquire function over time.

In this study, we investigated sequential ERV colonization events that have resulted in distinct lineages of retroviruses in the mule deer genome, none of which are fixed. Therefore, we are able to determine whether ERVs are silenced in the genome upon integration or if there is evidence for ERV activity that could affect host genome evolution. There have been four colonization events from retrovirus epizootics in mule deer since speciation with white tailed deer about one million years ago. Each lineage shows evidence of activity represented by extensive recombination within and among lineages. The youngest lineage is the only one with non-recombinant ERVs, which form several phylogenetic groups indicative of a new wave of infectious germline integrations. These data indicate that ERVs that colonized at different times during the relatively short evolutionary history of mule deer were transcriptionally active in the host genome and have generated new ERVs through recombination and that ERVs may enhance non-allelic homologous recombination events in the host genome. Our data also show that about 30% of ERVs have integrated near genes, and that some of these genes are now expressed from the ERV promoter. These data demonstrate that the colonization period of a new ERV family in mule deer is characterized by extensive ERV activity with the potential to affect both genome structure and gene regulation differentially among host populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-331

Genes involved in internalizing pathogens in Drosophila are shaped by recent and recurrent positive selection

Joo Hyun Im [1 2,*], Brian Lazzaro [2 3]

[1]Molecular Biology and Genetics, [2]Cornell Institute of Host-Microbe Interactions and Disease (CIHMID), [3]Entomology, Cornell University, Ithaca, United States

Abstract: Autophagy and phagocytosis are cellular mechanisms that internalize and eliminate intracellular and extracellular pathogens. The dynamic conflict between the host and pathogens that evolve to escape, resist or compromise host immunity can result in co-evolution, leading to a recurrent positive directional selection on host genes. We hypothesized that host phagocytosis and autophagy genes may experience such co-evolutionary pressure and therefore may show molecular evidence of adaptation. We performed population genetic analyses on phagocytosis and autophagy genes, as well as matching control genes, using previously published *Drosophila melanogaster* and *D. simulans* genome sequences. First, we detected a strong, recent selection on the genes involved in expansion of autophagosome and recognition and internalization of extracellular pathogens. For instance, we see a strong signature of selection at the *Atg8a* gene. *Atg8a* encodes a protein whose human homolog is known to be a target of an inhibitory effector in *Legionella pneumophila*. Nevertheless, we see distinct gene sets showing evidence of recent adaptation in *D. melanogaster* versus *D. simulans*, indicating that these two species may have faced unique challenges. Next, we observed several cases of an adaptive evolution in phagocytosis genes involved in particle recognition and degradation in both species. Although we see evidence of recent adaptation in individual genes, we do not find evidence that recent or recurrent positive selection is pervasive throughout entire functional classes of genes, which the exception of genes involved in degradation of phagocytized pathogens.

Expanded summary*: Host-pathogen interactions can result in various evolutionary patterns in the host ranging from positive directional selection to balancing selection. Previously, most population genetic studies on innate immunity have focused on the humoral immune response genes and the phagocytosis receptor genes, yet the evolution and the function of non-receptor phagocytosis and autophagy genes have not been extensively examined. Therefore, we surveyed the evolutionary models of host-pathogen interactions in the autophagy and phagocytosis genes and hypothesized that these internalization genes may have signatures of selection driven by the host-pathogen interactions. Understanding the evolutionary patterns of these genes provides us with a novel insight into the evolution of the innate immune responses and the clues on the biological function of these genes, offering the base to explore the host-pathogen interactions at a population scale.

To test the hypothesis, we first curated a set of phagocytosis and autophagy genes and other canonical immune genes, as well as control genes that are matched with target genes by gene length and genome location. We then calculated summary statistics of each gene from previously published sequences of 197 *D. melanogaster* from the Drosophila Genome Nexus Project and 20 *D.simulans* from Rogers et al (2014).

First, we looked at the summary statistics of individual genes and found a strong, recent selection on the genes involved in expansion of autophagosome. For instance, Atg8a exhibited a signal of recent selection in *D.melanogaster*. This is intriguing since the previous work showed that its proper conjugation with other host factors for autophagosome expansion in a human cell line is blocked by an effector protein from the intracellular pathogen *Legionella pneumophila*, supporting our hypothesis that Atg8a may be a target of host-pathogen coevolution. The ref(2)P, another case of recent selection in *D.melanogaster*, has already been known in *D.melanogaster* to be under selection due to the interaction with the sigma virus, adding to the rigor of this study. Furthermore, we detected an evidence of recent selection in phagocytosis genes that recognize and internalize extracellular pathogens. For example, crq, a recently characterized phagocytic receptor, is under positive selection only in *D.melanogaster*. The cases of recent selection are distinct in each species, suggesting that the two species may have experienced different selective pressures. In addition, we saw several cases of adaptive evolution in the phagocytosis genes that recognize and degrade foreign particles. For instance, Rbsn-5, a Rab5-binding effector responsible for phagosome maturation seems to be rapidly diverging between the species.

Next, we compared molecular evolutionary statistics of the genes in each functional class to those of the respective set of control genes to find out whether a particular functional class of these internalization pathways is under selection. We only found a group of genes that belong to the degradation class in phagocytosis in *D.melanogaster* to be evolving non-neutrally compared to the control

genes. Autophagy and phagocytosis pathways, as a group, did not exhibit the evidence of selection compared to the respective control genes, indicating that genes have experienced the selective pressure only at an individual level.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-334

Radiation of sugar transporters in the plant pathogenic genus Microbotryum

Sebastian Klenner [1,*], Ricardo C. Rodriguez de la Vega [2], Michael H. Perlin [3], Tatiana Giraud [2], Dominik Begerow [1]

[1]AG Geobotany, Ruhr- Universität Bochum, Bochum, Germany, [2]Ecologie, Systematique et Evolution, Universite Paris-Sud, Orsay, France, [3]Department of Biology, University of Louisville, Louisville, KY, United States

Abstract: Fungi are major agents of plant diseases and biotrophic pathogens are often characterized by intimate interactions with a host, often displaying a high degree of host specificity. A long lasting biotrophic phase requires well-adapted mechanisms to survive in the plant and efficient mechanisms for nutrient uptake are needed. While genes coding for cell-wall degrading enzymes are depleted in the genome of *Microbotryum lychnidis-dioicae*, the pathogen of *Silene latifolia*, the genes of various sugar transporter families seem to be expanded.

Phylogenetic analyses of a sugar transporter gene family of the genus *Microbotryum* provide insights into the evolution of sugar transporters. The questions of our study are 1) Does the phylogeny of sugar transporter genes reflect the phylogeny of the genus *Microbotryum*? 2) How dynamic is the family of putative sugar transporter genes within *Microbotryum*? 3) Are the various members of the gene family characterized by different mutation rates and selection regimes?

Our results indicate that upregulated sugar transporter genes are multiplied within the plant pathogenic genus *Microbotryum*. While some putative orthologs were detected in *Microbotryum* species infecting non-caryophyllacean host plants, preliminary data indicate a few members of the sugar transporter gene family to be unique to *Microbotryum* anther smuts of Caryophyllaceae. The potential impact of sugar transporter genes on adaptation processes and specialisation on respective host plants as well as their role in biotrophic interactions will be discussed based on phylogenetic analyses.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

OTH-HP2

Worlds within worlds: Scaling from molecules to populations in co-evolutionary dynamics

Andrea L. Graham [1,*]

[1]Ecology & Evolutionary Biology, Princeton University, Princeton, NJ, United States

Abstract: Why do hosts vary so much in how vigorously, and how effectively, they combat parasites? Reciprocally, why do parasites vary so much in how they exploit hosts? Robust empirical answers to these central questions in the evolutionary ecology of host-parasite interactions often focus on a single biological scale (e.g., that of the molecule, cell, organism or population). Theory demonstrates, however, that mechanistic, predictive understanding of co-evolutionary dynamics requires cross-scale analysis. For example, reciprocal immunoepidemiological feedbacks – i.e., between molecular heterogeneity among hosts and population-scale transmission of parasites – are hypothesized to maintain variation in host defenses. Empirical work to test such cross-scale hypotheses poses great challenges. Studies of the cross-scale causes and consequences of host and parasite heterogeneity in natural, parasite-rich environments are especially rare. It is rarer still for studies to take a detailed, long-term view: e.g., longitudinal measures of parasite burden and molecular and cellular immune parameters in individual mammals of known genotype, under well-characterized environmental variation. The collaborative study of the Soay sheep (Ovis aries) of St. Kilda, Scotland, and their gastrointestinal nematodes (e.g., Teladorsagia circumcincta) meets these restrictive criteria. Indeed, this system is proving suitable for crossing scales in co-evolutionary investigation. I will outline progress in understanding sheep-nematode co-evolutionary dynamics and parallels in experimentally-tractable systems.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

OTH-HP6

Molecular evolution of the Zika virus (ZIKV; Flaviviridae): role of RNA editing

Helen Piontkivska [1,*], Madeline Frederick [1], Marta Wayne [2], Michael Miyamoto [2]

[1]Biological Sciences, Kent State University, Kent, [2]Biology, University of Florida, Gainesville, United States

Abstract: Zika virus (ZIKV) is a mosquito-transmitted flavivirus that has been linked to microcephaly and fetal death in humans. We examined whether host-mediated RNA editing of adenosines (ADAR) plays a role in the molecular evolution of ZIKV. Using complete coding sequences for the ZIKV polyprotein, we show that potential ADAR substitutions are underrepresented at the ADAR-resistant GA dinucleotides of both the positive and negative strands, that these changes are spatially and temporally clustered (as expected of ADAR editing) for certain evolutionary lineages, and that ADAR mutagenesis can be linked to viral codon usage. Furthermore, resistant GA dinucleotides are enriched on the positive (but not negative) strand, which indicates that the former is under stronger purifying selection than the latter. ADAR editing also affects the evolution of the rhabdovirus Sigma (DMelSV). Our study now documents that host ADAR editing is a mutation and evolutionary force of positive- as well as negative-strand RNA viruses.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Host-parasite coevolution

POB-335

Unprecedented Eukaryotic Gut Microbiome Diversity within Long-Tailed Macaques (Macaca fascicularis) in Southeast Asia

Hope Hollocher*, Justin Wilcox [1]

[1]Biological Sciences, University of Notre Dame, Notre Dame, United States

Abstract: The majority of eukaryotes have been suggested to live on or in other hosts, but the diversity and ecology of these symbiotic eukaryotes remains consummately uncharacterized, despite unprecedented contemporary interest in prokaryotic microbiomes. Key ecological roles played by eukaryotes in free-living systems and the ubiquity of parasitism, commensalism, and mutualism in eukaryotes suggest that symbiotic eukaryotes may make important contributions to host-associated communities. While previous studies on the host-associated eukaryotic communities of vertebrates have reported low levels of diversity relative to both sympatric prokaryotic and free-living eukaryotic communities, these findings may be more indicative of differences in the methodologies used to characterize these communities than they are of actual ecological differences between these biological systems. To assess the potential for such hidden diversity within guts of non-human primates, we utilize a novel Illumina sequencing approach to characterize eukaryotic diversity within the feces of wild long-tailed macaques (*Macaca fascicularis*) on two islands in southeast Asia: Singapore and Bali, Indonesia. We report substantially higher levels of eukaryotic diversity than previously reported from the feces of primates. All five super-groups of eukaryotic life were represented, and several taxonomic groups were found to be common across all samples, suggesting the existence of a core eukaryotic community with the capacity to perform consistent ecological functions within these macaque hosts. Despite these commonalities, differences in eukaryotic gut assemblages were also detected that could be attributed to differences in host geography and diet. Our results are discussed within the context of how ecological guilds operating in the gut of macaques can drive community assemblage of symbiotic eukaryotes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

POB-97

Regulatory network inference from a dense time-course RNA-seq study of Drosophila innate immune response

M. Florencia Schlamp [1,*], Angela Early [1], Sumanta Basu [2], Andrew Clark [1 2]

[1]Molecular Biology and Genetics, [2]Biological Statistics and Computational Biology, Cornell University, Ithaca, United States

Abstract: Following microbial infection, *Drosophila* launch a rapid and efficient immune response that is crucial to survival. However, these responses are costly for the organism, consuming energy and resources that could be used for other life processes such as metabolism, reproduction, and environmental stress responses. This makes it advantageous for the organism to shut down the immune response as quickly as possible once the infection threat is passed. Organisms must tune their immune response to strike a balance between the advantage of a rapid and robust immune response to fight infection and the costly side-effects of an over-prolonged or unnecessary immune response. In this study, we perform a dense time-course RNA-seq analysis of the *Drosophila* immune response to learn more about the dynamics of activation and shutdown of the innate immune response. Flies were injected with commercial lipopolysaccharide (LPS), a known non-pathogenic elicitor that can stimulate a robust yet transient immune response while avoiding the confounding effects from a growing and changing internal population of pathogens. Flies were sampled for RNA-seq analysis pre-infection as a control and post-infection for 20 time points throughout 5 days. We used gene-wise linear models to fit polynomial trends with time, and standard empirical Bayes F-tests to select genes whose expressions altered significantly across the time course. We constructed networks of bivariate and multivariate Granger causality (GC) relationships among this subset of differentially expressed genes. GC relationships present in these networks point to several novel interactions governing temporal gene regulation of immune response.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

OW-II4

Evolutionary modification of the VLR-based adaptive immune system in jawless vertebrates

Sabyasachi Das [1,*], Jianxu Li [1], Yoichi Sutoh [1], Masayuki Hirano [1], Max cooper [1]

[1]Pathology, Emory University, Atlanta, United States

Abstract: Evolutionary modification of the VLR-based adaptive immune system in jawless vertebrates

Sabyasachi Das, Jianxu Li, Yoichi Sutoh, Masayuki Hirano, Max D. Cooper

Department of Pathology and Laboratory Medicine, Emory Vaccine Center, Emory University, Atlanta, GA 30322, USA

The three types of variable lymphocyte receptor genes, *VLRA*, *VLRB* and *VLRC*, in the extant jawless vertebrates, encode antigen receptors, the remarkably diverse repertoire of which is generated by insertion of neighboring leucine rich repeat (LRR) sequences into the incomplete germline genes. In both lampreys and hagfish, the B-cell like VLRB+ cells differentiate into VLRB-secreting plasma cells, whereas the αβ and γδ T cell-like VLRA+ and VLRC+ cells express their VLR products solely as cell surface proteins. However, in comparative studies of lampreys and hagfish, we find that remarkable functional differences have evolved in these lymphocyte lineages. In contrast with their striking predominance in lampreys, the VLRB+ cells constitute a minority lymphocyte population in hagfish, wherein VLRC+ cells instead predominate. Whereas the germline *VLRB* gene in hagfish contains a short non-coding intervening sequence, *VLRB* genes in sea lampreys and Japanese lampreys have very long intervening sequences containing many transposable elements, which may influence *VLRB* expression. In keeping with the relative low numbers of hagfish VLRB+ cells, we find that antibody responses to the same immunogen, sheep erythrocytes, are much less robust in hagfish than in lampreys. Remarkable differences are also found for the expression of non-VLR genes by the different lymphocyte populations in hagfish versus lampreys. Thus, even though the fundamental genetic program for differentiation of two prototypic T-like lymphocyte lineages and one B-like lineage is conserved in both jawless and jawed vertebrates, and therefore must have been present in a common vertebrate ancestor, the genetic programs used for fine tuning of the differentiation of these lymphocyte lineages have undergone notable independent evolutionary changes in lampreys and hagfish over the past ~480 million years.

Keywords: Immunity, Variable lymphocyte receptor, B cell, T cell, Immune system genes

Acknowledgements: This work was supported by National Institutes of Health Grants (R01AI072435, R01GM100151, and GM108838) and the Georgia Research Alliance.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

POB-94

A machine learning approach to predicting targets of the Dengue virus protease within the host proteome

Alex Stabell [1], Alison Gilchrist [1,*], Sara Sawyer [1]

[1]CU Boulder, Boulder, United States

Abstract: Dengue virus translates its genome into a single polyprotein and uses a self-encoded protease to cleave the polyprotein into individual proteins. The target site specificity for this protease is loose, so we anticipate that it also cleaves some host proteins in a manner that may be important for the viral life cycle. Using a machine learning approach, we have created a predictor to search for Dengue protease cleavage sites within the human proteome. We have experimentally validated that several human proteins in this candidate list are cleaved by the Dengue protease, and are important for replication. We are now examining each of these host-virus interactions to determine which might be acting as species-barriers in the movement of viruses between primate species. We are determining which of these genes are evolving rapidly in primate populations and experimentally testing different orthologs in Dengue infection assays. This project demonstrates how analysis tools from the field of molecular evolution can help facilitate novel characterization of host-virus biology.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

POB-95

How a fish lost its worm: natural tapeworm prevalence in threespine stickleback linked to heritable immune variation and parasite growth

Jesse Weber [1,*], Natalie Steinel [2 3], Brian Lohman [3], Daniel Bolnick [3]

[1]Division of Biological Sciences, University of Montana, Missoula, [2]Dell Medical School, [3]Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: Why might some populations experience no infections across years but have high-infection neighbors? Using a common garden approach, we tested whether threespine stickleback (*Gasterosteus aculeatus*) exhibit heritable variation in innate immune activity that could mediate natural differences in tapeworm (*Schistocephalus solidus*) prevalence. We exposed lab-reared stickleback from zero- and high-infection populations (as well as hybrid fish) to tapeworms, and subsequently measured gross infection phenotypes and several innate immune parameters. We observed large, heritable differences in immunity that were correlated with natural infection levels. Although stickleback from the zero-infection lake did become infected, they dramatically reduced tapeworm growth, enclosed and killed parasites in granulomas, and their granulocytes produced 3X-more reactive oxygen species (ROS) than cells from high-infection individuals. However, these immune responses were not without costs. ROS production was dysreguted in F2-generation hybrids—many of these fish produced more ROS than the zero-infection parents, and also suffered from perotinitis and excessive fibrosis when infected. Interestingly, fibrotic responses were absent from the high-infection fish, even when directly injected with immunogen and adjuvant, suggesting that some populations evolved to down-regulate immune responses and thereby tolerate infection. We are now using gene expression and QTL data to better understand the genetic basis behind these complex patters of host immune evolution.

Expanded summary*: Reciprocal selection pressures between hosts and parasites can drive rapid evolution, and geographically separated populations often follow distinct coevolutionary trajectories. This variation provides an excellent model for studying the evolution, genetics, and ecology of complex immune responses. As a postdoc at the U. of Texas I studied interactions between threespine stickleback and a tapeworm (*Schistocephalus solidus*). Tapeworm infections can greatly decrease host fitness, but immune responses are energetically costly and often self-damaging. I therefore predicted that stickleback would evolve a continuum of immune responses, ranging from resistance to tolerance, to optimize the benefits and costs of immunity in different environments. To test this prediction, I found host populations with large and likely adaptive differences in tapeworm infection prevalence, reared these fish and several parasite populations in the lab, and performed controlled exposures. The next few paragraphs describe results that I hope to present at SMBE, and which set a foundation for my future research projects.

Flow cytometry assays on granulocytes revealed that fish from a zero-infection lake (ZI) produce 3X higher levels of reactive oxygen species (ROS−toxic molecules used for both cell-signaling and attacking macroparasites) than fish from a high-infection lake (HI). Tapeworms were also *68-fold* smaller in the ZI fish, suggesting that these hosts resist infection by suppressing growth. Both traits varied continuously across >750 F2-generation hyrbid fish, indicating that they have polygenic architectures. Interestingly, ~20% of the hybrids produced higher ROS levels than even parental ZI animals. These transgressive phenotypes suggest that ZI fish evolved to both up- and down-regulate ROS levels, which comports with other data showing that ROS-responses: 1) consume energy, 2) damage host tissue, and 3) ZI fish evolved increased expression of ROS detox genes. In contrast, HI fish appear to tolerate infection; many infected hybrids and RL fish had severe organ fibrosis, but HI fish never showed this response, even when injected with antigens and adjuvant. Finally, hybrids and ZI animals often formed granulomas around tapeworms. Small worms in granumolas were often dead and digested, medium-sized worms were alive and only slightly damaged (large worms were never enclosed), and unenclosed worms showed no signs of damage.

Together, these data suggest that stickleback can evolve to both resist and tolerate a tapeworm parasite. There also appears to be functional interdependence between resistance traits, which may indicate the order in which they evolved: first worm growth is suppressed, then a granuloma response allows fish to enclose small worms, and finally increased ROS levels enable fish to attack and kill enclosed parasites without suffering excessive self-harm. This represents a rare opportunity to evaluate how multiple, modular immune responses arise and are integrated during evolution. Determining the molecular basis of each phenotype will enable tests of

this hypothesis. My collaborators and I have recently identified QTLs for all of these traits, as well used transcriptomic approaches to identify genes that are differentially expressed between ZI and HI fish. Several of the differentially genes have immune associations and lie within QTLs, making them excellent candidates for further study.
My future work will integrate these existing data with several other approaches, including experiments that test the consequences of resistance in nature, to further illuminate mechanisms and mutations that underlie stickleback immune divergence. This work has value not only for immunologists, evolutationary biologists, and geneticists, but is also likely to have medical applications. Several traits that I study are prominent features of disease (e.g., fibrosis), helminths infect a third of humanity and inflict serious damage on livestock, and understanding how hosts naturally counter or tolerate infection could lead to novel therapies.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

POB-96

RNA viruses drove adaptive introgressions between Neanderthals and modern humans

David Enard*, Dmitri Petrov [1]

[1]Stanford University, Stanford, United States

Abstract: Neanderthals and modern humans have interbred at least twice in the past 100,000 years and likely infected each other with viruses on both occasions. Here, we find that viruses drove an excess of tens of adaptive introgressions between Neanderthals and modern humans. This number is large enough that we were able to identify RNA viruses as one of the main drivers of adaptive introgressions from Neanderthals to European modern humans. Our results show that genetic variation in host genomes can be used to study ancient viral epidemics, with potentially important clues for the understanding of current and future epidemics.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

POA-416

QUANTIFYING HOST RESPONSES AND VIRAL EVOLUTION DURING ACUTE POLIOVIRUS INFECTION

Patrick Dolan*, Yinghong Xiao [1], Raul Andino [1]

[1]University of California, San Francisco, San Francisco, United States

Poster: The host environment encountered by pathogens *in vivo* is heterogeneous and dynamic. Individual tissues and cell types differ in their cellular and physicochemical composition and in their responses to innate immune signals. This heterogeneity creates distinct selective environments within the host. RNA viruses exhibit high mutation and recombination rates, existing as quasispecies, diverse collections of low-frequency mutants that surround the master genotype. How the complex selective environment *in vivo* shapes the viral population structure is largely unknown, but is likely to have important implications for tissue tropism, pathogenesis, transmission and treatment. Here, we trace the fate of poliovirus populations in infected mice to understand how tissue-specific selection drives changes in the composition of the viral quasispecies. In response to infection, we find that infected tissues deploy distinct antiviral transcriptional programs. These distinct responses correlate with the emergence of tissue-specific patterns of population diversity. This adaptation occurs without changing the consensus genotype, suggesting that the master genotype has been optimized by selection to maintain rapid access to functional subpopulations. When the evolutionary capacity of the virus population is inhibited, through the introduction of mutations in the viral polymerase that impair mutation and recombination, the viral population is unable to establish these tissue-specific population structures. This decreased adaptive capacity in these virus populations is correlated with attenuated phenotypes *in vivo*, suggesting that the ability of the population to rapidly establish specific population structures enables the virus to overcome tissue-specific barriers to infection and constitutes an important determinant of pathogenesis.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

POB-100

Early evolution of endogenous danger signal recognition by Toll-like receptor 4

Andrea Loes [1 2,*], Michael Harms [1 2]

[1]Institute of Molecular Biology, [2]Department of Chemistry and Biochemistry, University of Oregon, Eugene, United States

Abstract: The vertebrate Toll-like receptors (TLRs) induce inflammation in response to both pathogen-associated molecules and damage-associated molecules produced by the host. One model for their evolution is that TLRs initially recognized pathogen-associated molecules, but were later co-opted to recognize host danger signals. One such receptor is Toll-like receptor 4 (TLR4). In humans, TLR4 responds to the pathogen-molecule lipopolysaccharide (LPS) as well as host proteins S100A8, S100A9, and S100A12 (known collectively as calgranulins). LPS activation of TLR4 evolved in the ancestor of amniotes. No ortholog of the calgranulins has been identified outside of mammals, thus calgranulin activation of TLR4 has been considered a later evolutionary innovation. To test this model, we performed phylogenetic and synteny analysis of the calgranulins. We found that bird/reptile-specific MRP-126 was likely co-orthologous to the mammalian calgranulin clade. Using *ex vivo* cell-culture assays, we measured activation of TLR4 by calgranulin/MRP-126 from a variety of amniotes. We established that this pro-inflammatory axis is not mammalian-specific, but is shared across amniotes. We determined the specificity of these interactions by studying heterologous activation, revealing a conserved activation mechanism between receptor-ligand pairs. We found epistasis between members of the TLR4 complex from different species, revealing lineage-specific coevolution between members of the TLR4 complex and their activators. This work reveals that the pathogen and host-induced activation of TLR4 evolved over a similar evolutionary interval and has been maintained since. Further, this suggests that TLR4 is a general danger sensor, rather than a pathogen sensor that later evolved to recognize host danger signals.

Expanded summary*: Toll-like receptor 4 (TLR4) is a critical receptor that plays roles in both beneficial and pathological inflammatory responses. Like many TLRs, it activates inflammation in response to molecules derived from microbes (PAMPs: Pathogen-Associated Molecular Patterns) as well as molecules synthesized by the host (DAMPs: Damage-Associated Molecular Patterns). This allows the immune system to respond to danger signals generated by both “non-self” and “self” triggers. The order in which “non-self” and “self” danger signal recognition evolved for TLRs is poorly understood. Were these originally PAMP-receptors which later evolved to amplify danger signals through recognition of host DAMPs? Or is damage recognition an ancient feature of this family? Answering this question requires an understanding of the evolutionary history of TLRs.

TLR4 recognizes the bacterial cell wall component lipopolysaccharide (LPS), making it a critical front-line component for recognition of Gram-negative bacteria by the innate immune system. TLR4 also responds to a variety of DAMPs, most notably the calgranulin proteins S100A8, S100A9 and S100A12. Calgranulin DAMP activity plays important roles in wound healing and vascular development, but can also lead to upregulation and amplification of the inflammatory response in arthritis, arteriosclerosis, and inflammatory bowel disease.

Understanding the molecular basis for PAMP and DAMP activation of TLR4 is of active interest for the treatment of inflammatory disorders, as the ability to independently modulate PAMP and DAMP activity could allow suppression of pathological DAMP inflammation independently of the pathogen response. Understanding the evolutionary history of TLR4 could reveal the core logic of this immune receptor and allow for rational manipulation of these two types activation schemes, not just TLR4, but for TLRs in general. By isolating the historical intervals in which DAMP and PAMP activation evolved, we can reveal the genetic basis for functional expansion.

TLR4 acquired the ability to respond to LPS in the ancestor of amniotes, through the evolution of its co-factors MD2 and CD14. The evolutionary origins of DAMP recognition by TLR4 are less clearly understood. Members of the calgranulin clade are only found in mammals, and calgranulin-DAMP activity has only been validated for placental mammals. To elucidate how the evolution of DAMP activation relates to the evolution of PAMP recognition by TLR4, we set out to determine the historical interval in which calgranulin activation of TLR4 evolved.

Through phylogenetic and syntenic analysis, we identified that the bird/reptile-specific MRP-126 was likely co-orthologous to the mammalian calgranulin clade. Using an *ex vivo* cell culture assay, we determined that chicken MRP-126 can activate an inflammatory cascade through chicken TLR4. This demonstrates that calgranulin DAMP inflammation through TLR4 evolved at least in the ancestor of amniotes. We also determined that the co-factors required for LPS signaling through TLR4 are required for calgranulin

activation in amniotes. As LPS-PAMP and calgranulin-DAMP activation require the same minimal complex components for activation, it appears that the full TLR4-MD2-CD14 complex evolved for signaling by both PAMP and DAMP activation in an early amniote and has been maintained since. Finally, we revealed a conserved activation mechanism between receptor-ligand pairs by assessing heterologous activation between species (e.g. human S100A9 on chicken TLR4). This indicates that the molecular basis for calgranulin-DAMP activity has been maintained across amniotes.
This work reveals that TLR4 was a general danger sensor in early amniotes, rather than a pathogen sensor that later evolved to recognize host danger signals. Expression of TLR4 has been observed non-amniote vertebrates which lack cofactors MD2 and CD14. Identification of ligands recognized by TLR4 in these species may illuminate the ancient role of TLR4 in vertebrate innate immunity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

POB-98

Immune-driven signals of selection in diverse human populations

Sarah Kaewert [1,*], Christina Eichstaedt [2 3], Luca Pagani [4], Evelyn Jagoda [5], Florian Clemente [6], Georgi Hudjashov [4], Tiago Antao [7], Toomas Kivisild [1 4]

[1]Department of Archaeology and Anthropology, University of Cambridge, Cambridge, United Kingdom, [2]Centre for Pulmonary Hypertension, Thoraxclinic at the University Hospital Heidelberg, [3]Institute of Human Genetics, Heidelberg University, Heidelberg, Germany, [4]Estonian Biocentre, Tartu, Estonia, [5]Department of Human Evolutionary Biology, Harvard University, Cambridge, United States, [6]Institut de Biologie Computationnelle, Université Montpellier, Montpellier, France, [7]Division of Biological Sciences, University of Montana, Missoula, United States

Abstract: Introduction: Immune-related genes have been shown in many studies to be convincing targets of past selection events. However, relatively few populations have been represented. By including underrepresented global populations in selection studies, we can improve our understanding of human genetic diversity as it relates to immunity, as well as examine human adaptation to immune challenges at a higher resolution.
Materials and Methods: Genomes from the Estonian Biocentre Human Genome Diversity Panel and other sources were compiled to create a dataset of 376 high coverage genomes representing thirteen geographically diverse regional groups. Using the results of 3 different statistics for positive selection, enrichment tests were performed for immune-related genes under possible selection. Additional filters were applied to find SNPs potentially driving signals of selection.
Results: Enrichment tests suggest that non-African populations are more likely to be significantly enriched in various immune phenotypes than African populations. Windows containing genes implicated in pathogen interaction were most likely to show significant enrichment. The list of SNPs resulting from additional filters represents potential drivers of selection in each of the thirteen populations, many of which are in genes with immune function.
Conclusions: Non-African populations show more evidence of selection in immune genes and pathogen interaction genes. Some of the strongest signals come from populations underrepresented in the genomic literature, such as Siberian and South American populations. This underlines the importance of using diverse population samples in population genetic studies, since otherwise this diversity can be missed.

Expanded summary*: As of 2010, around 14% of the human genome has been reported to be under positive selection by at least one study[1]. Of the variety of functional classes of genes found to be under positive selection[2], immune-related genes have been shown in multiple studies to exhibit some of the strongest signals of selection[1, 3, 4, 5]. Not only do some immune-related genes display strong signals of selection, but the immune gene class as a whole is significantly overrepresented in regions found to have experienced recent positive selection[1, 6, 7].
Balancing selection is just as important as positive selection when relating to immunity, if not more so. Though different studies have found different candidates for genes under balancing selection, immunity is consistently one of the categories found to be significant[8], and is overrepresented as a category in the lists of candidate genes[9]. Additionally, though many signatures of balancing selection are shared across populations, many of those that are unique are found in immune-related genes[8].
While the importance of selection on immune genes and its role in human evolution has been established, the diversity of populations represented in selection studies has been relatively small until recently. Because of this, there is a significant gap in the literature regarding selection on immune genes in various underrepresented populations. A study of selection signals in these populations should yield new insights into their population history and shed light on the challenges faced by humans as they expanded across the globe.
This study uses 376 high coverage whole genome sequences from the Estonian Biocentre Human Genome Diversity Panel to form 13 population groups based on geography. Top results from multiple window-based selection statistics for positive and balancing selection were analyzed for enrichment of immune-related genes, as well as filtered for predicted functional importance in order to find potential driver SNPs. The top 1% of windows in each population and test were analyzed for enrichment in immune-related

genes, with many populations showing significant enrichment in multiple tests, especially in genes related to pathogen interaction. Interestingly, non-African populations showed the most enrichment, suggesting adaptation as populations migrated out of Africa and beyond. Additionally, some of the strongest signals come from populations that have rarely been represented in such selection scans before, such as Siberian and South American populations. In including more genetic diversity in selection studies, we can potentially see selection events that have been missing from earlier studies. With more examples, we can also learn more about immune-driven adaptation in humans.

1. Barreiro, L. B. and Quintana-Murci, L. (2010). From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nature Reviews Genetics*, 11(1):17–30.
2. Akey, J. M. (2009). Constructing genomic maps of positive selection in humans: Where do we go from here? *Genome Research*, 19(5):711–722.
3. Karlsson, E. K., Kwiatkowski, D. P., and Sabeti, P. C. (2014). Natural selection and infectious disease in human populations. *Nature Reviews Genetics*, 15(6):379–393.
4. Quintana-Murci, L. and Clark, A. G. (2013). Population genetic tools for dissecting innate immunity in humans. *Nature Reviews Immunology*, 13(4):280–293.
5. Siddle, K. J. and Quintana-Murci, L. (2014). The Red Queen's long race: human adaptation to pathogen pressure. *Current Opinion in Genetics & Development*, 29:31–38.
6. Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A Map of Recent Positive Selection in the Human Genome. *PLoS Biol*, 4(3):e72.
7. Wang, E. T., Kodama, G., Baldi, P., and Moyzis, R. K. (2006). Global landscape of recent inferred Darwinian selection for Homo sapiens. *Proceedings of the National Academy of Sciences of the United States of America*, 103(1):135–140.
8. Key, F. M., Teixeira, J. C., de Filippo, C., and Andrés, A. M. (2014). Advantageous diversity maintained by balancing selection in humans. *Current Opinion in Genetics & Development*, 29:45–51.
9. Andrés, A. M., Hubisz, M. J., Indap, A., Torgerson, D. G., Degenhardt, J. D., Boyko, A. R., Gutenkunst, R. N., White, T. J., Green, E. D., Bustamante, C. D., Clark, A. G., and Nielsen, R. (2009). Targets of Balancing Selection in the Human Genome. *Molecular Biology and Evolution*, 26(12):2755–2764.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

OW-II6

Genetic variation in social interactions leads to outbreaks of a sexually-transmitted fungus in D. melanogaster

C. Nick Keiser [1], Julia Saltz*

[1]Rice University, Houston, United States

Abstract: When will a single index case turn into an outbreak, or even an epidemic? Because many infectious diseases, especially sexually-transmitted infections, are spread via social contacts, understanding how individual behaviors lead to group-wide incidence of infections is critical for predicting disease spread. Although the importance of social interaction patterns is widely recognized in disease ecology, less is known about *variation* in the behavioral causes of pathogen transmission across host genotypes and how these behaviors are affected by their social group.

Here, we describe a series of experiments investigating the interplay between sexual behaviors and pathogen transmission in social groups of *D. melanogaster*. We exposed male or female "index case" flies of three different genotypes to a generalist entomopathogenic fungus that is transmitted horizontally, especially during mating, and then placed them into social groups that varied in sex ratio. By measuring who mated with whom, and tracking the spread the fungus, we find that (i) genotypes differ in susceptibility; (ii) the overall incidence of disease changed with group sex ratio; and (iii) unexpectedly, disease outbreaks were larger when the index case was female, despite females' low mating rates. Our findings that genetic and behavioral variation among individuals and groups jointly determine groups' vulnerability to outbreaks, and suggest that selection on immune systems may depend on social context and vice versa.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

OW-II5

Using Time-Series Sequence Data to Characterize the Interaction Between the Host Antibody Repertoire and Pathogens

Nicolas Strauli [1,*], Satish Pillai [2 3], Ryan Hernandez [4 5 6]

[1]Biomedical Sciences, [2]Department of Laboratory Medicine, University of California at San Francisco, [3]Blood Systems Research Institute, [4]Institute for Quantitative Biosciences (QB3), [5]Institute for Human Genetics, [6]Department of Bioengineering and Therapeutic Sciences, University of California at San Francisco, San Francisco, United States

Abstract: The ability to identify the antibodies (Abs) that are interacting with a given pathogen in humans is of immense importance for the fields of drug development, vaccine development, and the broader discipline of immunology. However, current laboratory techniques for executing this are low throughput, typically yielding only a handful of Abs that bind a given antigen. Given that there is likely a plethora of Ab lineages that respond to an infection within an individual, we seek to globally identify those that make up this response, and in turn, characterize their interaction with a given pathogen. We accomplish this by leveraging longitudinal deep sequencing data of individuals' antibody repertoires (AbRs), as well as pathogenic populations. Deep sequencing data of the AbR provide a snapshot of the frequencies of Ab lineages in an individual. By deeply sequencing the AbR over many time-points, one can observe how these Ab lineage frequencies change over time. Just as population geneticists can use the frequency trajectories of alleles over time to identify those that are under selection, we can use the frequency trajectories of Ab lineages to identify those that are responding to a given antigen. We first validate this approach by applying it to a dataset generated by Henn *et al.* 2013[1], in which 5 individuals were vaccinated with the same tri-valent influenza vaccine (TIV), and RNA-sequencing (RNAseq) was performed on their peripheral blood mononuclear cells for 11 consecutive days. We developed a bioinformatic pipeline to extract the AbR from their RNAseq data, and then use statistical methods based in the discipline of functional data analysis (FDA)[2] to identify the Ab lineages that are likely responding to TIV. We compare TIV-responding Abs identified in our study to demonstrated influenza-targeting Abs in the literature, and find that they are similar. We then seek to apply this approach to serial blood samples from individuals infected with HIV. We have deep sequenced the HIV population as well as the AbR in 10 untreated HIV-infected individuals, at 10-20 time-points each, spanning 4-13 years. To identify the HIV and Ab lineages that are interacting over time, we again use FDA-based methods to find the pairs of lineages that significantly co-vary. We expect that this will yield a systems level view of the AbR/HIV interaction in these patients.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

POB-104

In-depth characterization of a hallmark for balancing selection: HLA heterozygote advantage against HIV-1

Jatin Arora [1,*], Federica Pierini [1], Paul Mclaren [2], Jacques Fellay [3], Tobias Lenz [1]

[1]Max Planck Institute for Evolutionary Biology, Plön, Germany, [2]JC Wilt Infectious Diseases Research Center, Winnipeg, Canada, [3]École Polytechnique Fédérale de Lausanne, Laussane, Switzerland

Abstract: Pathogen-mediated balancing selection may drive host immunogenetic diversity. A hallmark for balancing selection in humans is heterozygote advantage at genes of the Human Leukocyte Antigen (HLA), resulting in durable HIV-1 control. However, the mechanism through which heterozygotes obtain an advantage is still elusive. It may be conferred by the ability of HLA heterozygotes to present more different viral peptides to immune cells, possibly resulting in more efficient cytotoxic T-cell responses. Heterozygosity may also simply increase the chance to carry the most protective HLA variants, as individual HLA alleles differ substantially in their association with HIV-1 control. Taking advantage of HLA genotype and set point viral load data from 6,311 HIV-1 patients of European ancestry, we find a lower viral load for heterozygotes at HLA-B ($P < 0.0001$) and HLA-C ($P = 0.022$). Screening the entire HIV-1 proteome, we observed that patients heterozygous at HLA-B and HLA-C are predicted to bind a broader array of HIV-1 epitopes ($P < 0.0001$ for both loci). Interestingly, a patient's viral load correlated negatively with the breadth of the patient's HLA-bound HIV-1 epitope repertoire for HLA-B (tau = -0.15, $P < 10^{-16}$), but not for HLA-C (tau = 0.01, $P = 0.09$), suggesting that heterozygote advantage at HLA-B is mediated by a quantitative cytotoxic T-cell response, but that different mechanisms could be involved at HLA-C. We also analyzed autologous HIV-1 sequence data and observed a significantly higher divergence of HIV-1 strains among HLA-B heterozygous patients compared to homozygotes ($P = 0.025$), suggesting stronger evolutionary pressure from HLA heterozygosity.

Expanded summary*: The genes of Human leukocyte antigen (HLA) locus code for the cell-surface molecules which present pathogen-derived peptides to immune cells. The locus is featured by exceptionally high allelic diversity and is regarded as a hallmark for balancing selection in humans. It has been associated with many infections, where HIV is among the most robust associates across multiple ethnicities. HLA alleles bind unique repertoires of peptides and are known to be differentially associated with HIV control. Fine mapping of HIV-1 associated HLA variation to the peptide-binding groove (McLaren et al. 2015) suggests a key role for HLA-presentation of HIV-1 epitopes in disease control. In our previous study, using HLA genotype and set point viral load (spVL) data of 6,311 HIV-1 patients of European ancestry, we computationally screened entire HIV-1 proteome for disease-associated epitopes and characterized a core set of HLA-bound HIV-1 epitopes that together accounted for the same amount of variation, 12.3%, in spVL as all previously associated independent genetic variants in HLA (McLaren et al. 2015). We observed a negative correlation between an HLA-B allele's effect on viral load and the number of HIV-1 peptides it was predicted to bind (tau = -0.26, $P = 0.002$), suggesting the quantitative advantage of HLA-presentation of broader HIV-1 epitope repertoires. However, epitope-specific association with spVL revealed that the epitopes bound by the same HLA allele did not necessarily have the same effect on viral load, highlighting qualitative aspect of HLA presented HIV-1 epitopes.

In the present study, we have analyzed these quantitative and qualitative aspects in-depth by taking heterozygote advantage, a conventional form of balancing selection, at HLA as the model. The actual mechanism behind HLA heterozygote advantage is still elusive. It may be conferred by the ability of HLA heterozygotes to present a broader array of pathogenic peptides to immune cells or by simply increasing the chance to carry more protective HLA variants. Using the same data, we found that heterozygosity at HLA-B and HLA-C was significant associated with lower viral load and (HLA-B $P < 0.0001$; HLA-C $P = 0.022$). Patients heterozygous at HLA-B and HLA-C were predicted to bind broader array of HIV-1 epitopes ($P < 0.0001$ for both loci). Interestingly, a patient's viral load was negatively correlated with the breadth of patient's HLA-bound HIV-1 epitopes for HLA-B (tau = -0.15, $P < 10^{-16}$), but not for HLA-C (tau = 0.01, $P = 0.09$), suggesting that heterozygote advantage at HLA-B is mediated by a quantitative CTL response, but a different mechanism could be involved at HLA-C. The observed significantly higher divergence of HIV-1 strains among HLA-B heterozygote patients compared to homozygotes ($P = 0.025$) suggests stronger evolutionary pressure exerted by HLA heterozygosity. Larger number of acquired polymorphisms in HLA heterozygotes than in homozygotes might result in higher cumulative fitness cost leading to stronger virus attenuation.

All together, these findings provide a proximate functional link for the robustly established association between HLA and HIV control. They suggest that both the quantity and the quality of HLA bound HIV epitopes are important for effective anti-viral immune response, favoring retention of multiple alleles of HLA in the population. Our approach of combining computational HLA-specific epitope prediction with disease phenotype validation may provide a promising avenue for identification and prioritization of novel disease-associated epitopes as potential therapeutic targets. These findings also lend the support to the vaccine programs that aim to impart anti-viral immunity by exposure to broad yet particular array of HIV peptides.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

OW-II1

How social status changes the immune system: experimental evidence from rhesus macaques

Jenny Tung [1,*], Noah Snyder-Mackler [1], Joaquin Sanz [2], Jordan Kohn [3], Jessica Brinkworth [4], Shauna Morrow [1], Amanda Shaver [1], Jean-Christophe Grenier [2], Roger Pique-Regi [5], Zachary Johnson [3], Mark Wilson [3], Luis Barreiro [2]

[1]Evolutionary Anthropology, Duke University, Durham, United States, [2]Genetics, Centre Hospitalier Universitaire Sainte-Justine Research Center, Montreal, Canada, [3]Yerkes National Primate Research Center, Emory University, Atlanta, [4]Anthropology, University of Illinois-Urbana-Champaign, Urbana, [5]Center for Molecular Medicine and Genetics, Wayne State University, Detroit, United States

Abstract: In hierarchically organized species, social status can strongly influence fertility, survival, and other fitness-related traits. To understand the molecular pathways that mediate these relationships, we used social status manipulations in female rhesus macaques to investigate the causal effects of dominance rank on immune gene regulation. We constructed 9 social groups (n=5 females per group) using a well-established paradigm in which order of introduction predicts dominance rank: earlier introduced animals attain higher status. Using high-throughput assays of gene regulation (RNA-seq and ATAC-seq), we show that (i) social status causally alters gene expression and chromatin accessibility in immune cells, often in a cell type-specific manner, and (ii) these effects are largely plastic with changes in rank. Social status also influences the response to infection. Using *ex vivo* challenge experiments with the bacterial compound lipopolysaccharide (LPS) and the viral mimetic Gardiquimod, we find that low ranking females respond more strongly to immune challenge and that rank predicts the relative use of the alternative arms of the Toll-like receptor 4 signaling pathway: low status polarizes TLR4 signaling towards a more pro-inflammatory response. Finally, *post hoc* analysis indicates that the topology of gene regulatory networks following immune challenge also systematically differs depending on dominance rank, highlighting gene pathways that differ from those identified in single gene analyses. Together, our results indicate a strong link between social behavior and the immune system, with important ramifications for understanding the molecular targets of social selection pressures and the evolution of social hierarchies more broadly.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

POB-102

Genome-wide methylation profiling of Wolbachia infection in a parasitic wasp

Xin Wu [1,*], Dan Sun [1], Amelia Lindsey [2], Paramita Chatterjee [1], Jack Werren [3], Richard Stouthamer [2], Soojin Yi [1]

[1]Georgia Institute of Technology, Atlanta, [2]University of California, Riverside, [3]University of Rochester, Rochester, United States

Abstract: Epigenetic modifications of genomic DNA and the packaging proteins (histones) have the potential to modulate molecular profiles of organisms in response to changes of environmental signals. Recent studies have demonstrated the potential roles of DNA methylation to influence insect phenotypes in the context of development, social roles, and diseases. We investigated how DNA methylation of the parasitic wasp *Trichogramma pretiosum* is altered due to intracellular parasitic bacterium *Wolbachia*. Specifically, in *Trichogramma* wasps, *Wolbachia* is transmitted vertically and induces parthenogenesis in females, a process in which unfertilized eggs develop into viable adult females. We analyzed whole-genome bisulfite sequencing maps of *Trichogramma* strains with and without *Wolbachia* infection. Hundreds of genes are differentially methylated between infected and noninfected strains, indicating substantial epigenetic changes accompanying *Wolbachia* infection. These differentially methylated genes are enriched in evolutionarily conserved hymenopteran genes, suggesting that *Wolbachia* infection targets a core set of genes and may involve a systematic pathway shared across taxa. Interestingly, we also identify numerous sites where DNA methylation changes are particularly drastic between the infected and noninfected strains. Characterizing epigenetic changes due to *Wolbachia* infection will enable us to gain insights into the mechanisms of parthenogenesis in insects and the consequences of a distorted gender ratio. Furthermore, understanding the epigenetic impact of *Wolbachia* infection can potentially improve the use of *Trichogramma* as a pest control agent.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

POB-108

Genetic underpinnings of sarcoptic mange susceptibility and severity in Yellowstone wolves (Canis lupus)

Alexandra DeCandia*, Emily Almberg, Daniel Stahler, Bridgett vonHoldt

Abstract: A classic paradigm in population genetics and species management posits that immunogenetic variation buffers against individual and population level disease risks. In its simplest form, this variation confers organisms and their populations with multiple immune responses, thereby limiting a pathogen's ability to evolve exploitation strategies of common weaknesses. As a result, numerous studies characterize variation in genes with known immune function, with few delving deeper into genomic analyses. Here, we seek to expand this paradigm and consider the role of genome-wide variation in governing individual disease state. Using sarcoptic mange in Yellowstone National Park (YNP) wolves as our study system, we survey immunogenetic variation and perform a genome-scale family-based association study to identify loci linked with disease susceptibility and severity. Critically, these analyses consider both traditional immune genes and loci typically excluded from more targeted immunogenetic approaches. Of equal importance, they are conducted in a natural population of reintroduced carnivores currently experiencing disease-mediated morbidity and mortality. Through their highly resolved pedigree and detailed longitudinal data, YNP wolves enable integration of environmental, demographic, and genomic parameters when evaluating risk factors of disease. The information gleaned from their study and integrative methods described herein may then inform management of similar reintroductions. We thus illustrate the importance of considering multiple facets of genomic variation in the complex ecology of wild vertebrates and disease, and hope others adopt this approach for better monitoring, management, and evolutionary inference going forward.

Expanded summary*: It is widely established that infectious disease poses a serious threat to endangered wildlife and the humans around them [1-3]. Particularly in species with small, fragmented, or reintroduced populations, the inability to cope with novel or enduring parasites can precipitate population declines and ultimate extinction. Maintaining high levels of genetic diversity at immune genes (such as the major histocompatibility complex or MHC gene family) is one way to buffer against these threats. It confers a population with multiple immune responses, thereby limiting a pathogen's ability to evolve exploitation strategies of common weaknesses [4,5]. In some cases, immune genes represent the only remaining source of genetic diversity in an otherwise monomorphic population [6]. As a result, numerous studies characterize MHC diversity in vertebrates [7-9], with few delving deeper into genome-wide analyses [10].

Here, we link genome-wide variation to disease state in a population of successfully reintroduced carnivores: Yellowstone National Park (YNP) wolves. This population is an ideal focal system for this study due to their history of close population monitoring and management, sample and data accessibility, and newly established disease load (particularly regarding sarcoptic mange). By leveraging their highly resolved pedigree [11] and fine scale phenotypic data, we explore the role of genetics in sarcoptic mange susceptibility and severity through the following three aims:

Aim One. Survey immunogenetic variation in YNP wolves by genotyping microsatellite loci in the MHC gene family. As previously mentioned, numerous studies characterize MHC diversity in vertebrates as a convenient and comparable proxy for immunogenetic variation [8,12-14]. We therefore begin our exploration of the genetics underlying mange susceptibility and severity with a broad scale population survey of immunogenetic variation in YNP wolves. By calculating diversity statistics across 24 MHC microsatellite loci, we estimate baseline levels of immunogenetic diversity and compare these levels to previously published neutral microsatellite variation (often used as a proxy for overall genetic diversity) [11,15,16].

Aim Two. Fit a linear model to identify significant predictors of mange infection. We then explore the relationship between demographic parameters (such as sex, age, coat color, pack, pack size, social status, and parent infection status), environmental variables (location in park, year, and season), genetic parameters (MHC microsatellite diversity, neutral microsatellite diversity, and inbreeding coefficient), and disease state (as measured by multiple mange metrics including presence/absence, most severe mange score, *etc.*) through regression analyses. This elucidates which parameters best predict mange infection and correlate with infection severity in YNP wolves.

Aim Three. Perform a family-based association study to identify genetic variants associated with mange infection at the genome-scale. Lastly, we expand these analyses to explore the role of genome-wide variation in mange infection and severity. Here,

we use the Yellowstone pedigree to perform a family-based association study with a dataset of single nucleotide polymorphisms (SNPs) generated by the Genotyping-by-Sequencing (GBS) method. This type of analysis tracks specific alleles through mange-exposed families, thereby identifying loci and nearby genes significantly correlated with disease state throughout the genome.

Each of these aims advances previous and ongoing research on population genetics [11,15,16] and disease [17-19] in YNP wolves [20]. By integrating multiple datasets, our investigation of genetic drivers of disease susceptibility and severity will provide novel insight into disease dynamics operating at the individual and population levels in YNP. Information gleaned from these analyses may then inform best practices at all stages of reintroduction for similar species management plans undertaken by governmental and non-governmental organizations [21]. More generally, this work contributes to the growing body of research characterizing immunogenetic diversity in wild populations of vertebrates. It is among the first to consider genome-wide diversity and associate specific loci with infectious disease risk in a wild population. This challenges the primacy of targeted immunogenetic approaches in molecular disease ecology and underscores the importance of considering multiple datasets (immunogenetic, genomic, environmental, demographic, *etc.*) in the monitoring, management, and evolutionary study of wild populations threatened by disease.

Citations

[1] Lyles, A. M. & Dobson, A. P. Infectious disease and intensive management: population dynamics, threatened hosts, and their parasites. *Am. Assoc. Zoo Vet.* **24,** 315–326 (1993).

[2] Murray, D. L., Kapke, C. a., Evermann, J. F. & Fuller, T. K. Infectious disease and the conservation of free-ranging large carnivores. *Anim. Conserv.* **2,** 241–254 (1999).

[3] Woodroffe, R. Managing disease threats to wild mammals. *Anim. Conserv.* 185–193 (1999).

[4] Hedrick, P. Balancing selection and MHC. *Genetica* **104,** 207–214 (1999).

[5] Bergstrom, C. T. & Antia, R. How do adaptive immune systems control pathogens while avoiding autoimmunity? *Trends Ecol. Evol.* **21,** 22–8 (2006).

[6] Aguilar, A. *et al.* High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 3490–4 (2004).

[7] Hedrick, P., Lee, R. & Parker, K. Major histocompatibility complex (MHC) variation in the endangered Mexican wolf and related canids. *Heredity (Edinb).* **85,** 617–624 (2000).

[8] Kennedy, L. J. *et al.* Major histocompatibility complex diversity in the endangered Ethiopian wolf (Canis simensis). *Tissue Antigens* **77,** 118–25 (2011).

[9] Lau, Q., Chow, N. & Gray, R. Diversity of MHC DQB and DRB Genes in the Endangered Australian Sea Lion (Neophoca cinerea). *J. Hered.* 395–402 (2015).

[10] Morris, K., Wright, B. & Grueber, C. Lack of genetic diversity across diverse immune genes in an endangered mammal, the Tasmanian devil (Sarcophilus harrisii). *Mol. Ecol.* **24,** 3860–72 (2015).

[11] vonHoldt, B. M. *et al.* The genealogy and genetic viability of reintroduced Yellowstone grey wolves. *Mol. Ecol.* **17,** 252–74 (2008).

[12] Hedrick, P., Lee, R. & Buchanan, C. Canine parvovirus enteritis, canine distemper, and major histocompatibility complex genetic variation in Mexican wolves. *J. Wildl. Dis.* **39,** 909–913 (2003).

[13] Arbanasić, H. *et al.* Extensive polymorphism and evidence of selection pressure on major histocompatibility complex DLA-DRB1, DQA1 and DQB1 class II genes in Croatian grey wolves. *Tissue Antigens* **81,** 19–27 (2013).

[14] Marshall, H. D., Langille, B. L., Hann, C. A. & Whitney, H. G. Patterns of MHC-DRB1 polymorphism in a post-glacial island canid, the Newfoundland red fox (Vulpes vulpes deletrix), suggest balancing selection at species and population timescales. *Immunogenetics* **68,** 381–389 (2016).

[15] Jankovic, I., vonHoldt, B. M. & Rosenberg, N. a. Heterozygosity of the Yellowstone wolves. *Mol. Ecol.* **19,** 3246–9 (2010).

[16] vonHoldt, B. & Stahler, D. A novel assessment of population structure and gene flow in grey wolf populations of the Northern Rocky Mountains of the United States. *Mol. Ecol.* **19,** 4412–27 (2010).

[17] Almberg, E. S., Mech, L. D., Smith, D. W., Sheldon, J. W. & Crabtree, R. L. A serological survey of infectious disease in Yellowstone National Park's canid community. *PLoS One* **4,** e7042 (2009).

[18] Almberg, E. S., Cross, P. C., Dobson, a. P., Smith, D. W. & Hudson, P. J. Parasite invasion following host reintroduction: a case study of Yellowstone's wolves. *Philos. Trans. R. Soc. B Biol. Sci.* **367,** 2840–2851 (2012).

[19] Almberg, E. S., Cross, P. C. & Smith, D. W. Community Persistence of canine distemper virus in the Greater carnivore Yellowstone community Ecosystem's. *Ecol. Appl.* **20,** 2058–2074 (2014).

[20] Smith, D. *et al. Yellowstone National Park Wolf Project Annual Report 2013*. (2014).

[21] Wayne, R. & Hedrick, P. Genetics and wolf conservation in the American West: lessons and challenges. *Heredity (Edinb).* **107,** 16–9 (2011).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

POB-106

Pervasive loss of mutability during the evolution of long-lived anti-HIV B cell lineages

Marcos Vieira [1,*], Sarah Cobey [1]

[1]Ecology and Evolution, University of Chicago, Chicago, United States

Abstract: Antibodies evolve in B cell populations under selection for improved antigen recognition. The nucleotide sequences of antibody genes regulate their own mutation and have likely evolved under selection for improved adaptability during immune responses. How the adaptability of antibodies changes as their sequences evolve, however, is unclear. The enzymes that mutate antibody genes preferentially target highly mutable 'hotspots', which help sustain high mutation rates across the sequence and help target mutations to regions where they are likely to increase antigen binding without destabilizing the protein. Since hotspots are highly likely to mutate, repeated hotspot losses during B cell evolution might limit subsequent antibody adaptation. Evidence for widespread hotspot losses during antibody evolution, however, is inconclusive. Other mechanisms, such as selection to retain mutability, might counteract the intrinsic propensity for hotspots losses. We performed a detailed reconstruction of the evolution of mutational hotspots in long-lived anti-HIV B cell lineages. By inferring changes in mutability between consecutive nodes in B cell genealogies, we show that mutability losses are widespread during antibody evolution and occur more frequently than gains. We found no evidence of selection to retain mutability. Simulations based on a previously developed 5-mer mutability model reproduced observed changes in other mutability metrics, suggesting the model accurately describes among-site heterogeneity in mutation rates during antibody evolution. Widespread mutability losses may constrain the long-term evolution of antibodies against chronic or repeated infections, such as influenza and HIV.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

POB-107

The fish adaptive immune response and its suppression by helminths

Natalie Steinel*, Jesse Weber, Daniel Bolnick

Abstract: Parasitic worm (helminth) infections can immunosuppress hosts, leaving them vulnerable to co-infection. This talk will present a model of helminth immunosuppression: the bony fish (teleost) stickleback (*Gasterosteus aculeatus*) and its helminth parasite, *Shistocephalus solidus*. The teleost adaptive immune response is thought to occur in the melano-macrophage center (MMC). This structure is the proposed site in which B cells proliferate and generate high affinity antibody, homologous to germinal centers in mammals. Stickleback MMCs transiently increase in size following immunization, suggesting that the MMC likely participates in the stickleback adaptive immune response. Infection of stickleback with *Shistocephalus solidus*, however, suppresses MMC size. This type of helminth-mediated immunosuppression has been observed in other species, including humans. To investigate the mechanisms underlying parasite-mediated immunosuppression in teleosts, wild populations of Vancouver Island stickleback were identified that vary in their magnitude of helminth-mediated immunoregulation. Helminth-infected stickleback from Gosling lake show suppressed MMC responses, while Roberts lake stickleback are refractory to helminth-induced immunosuppression. The resistance of Roberts lake fish to immunosuppression, however, is helminth genotype specific, with some helminth genotypes able to suppress MMC size. F2 hybrid fish were generated from Gosling and Roberts populations and QTL analysis was performed to identify loci associated with differential immunoregulation of the MMC response by helminths. As MMCs are found in most teleosts (as wells as some reptiles and amphibians), these findings highlight a new and broadly applicable assay of immunity, which does not require the development of species-specific reagents. This approach has the potential to be used in the study of teleost host-pathogen interactions both in the laboratory and in the wild.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

OW-II2

The immune system was a major target of natural selection during the European Neolithic

Iain Mathieson [1,*]

[1]Genetics, Harvard Medical School, Boston MA 02115, United States

Abstract: The European Neolithic, starting around 6,400 BCE, marked a dramatic change in terms of both population and lifestyle. Over several thousand years hunter-gatherer populations merged with migrating farmers who spread their settled agricultural lifestyle across the continent. Improvements in the efficiency of ancient DNA (aDNA) sequencing mean that aDNA from this period can be used, not only to track population movements, but also to detect natural selection – revealing how these populations adapted to changes in environment, diet, and social organization.

We analyze data generated in the Reich lab from more than 300 individuals who lived between 10,000 and 1,500 BCE and show that the immune system, along with diet- and pigmentation-related traits, was a major target of natural selection. Several genes involved in pathogen response were targeted, including haplotypes at the *OAS* and *TLR* gene clusters that had originally introgressed from Neanderthals. We detect at least six independent targets of selection at the Major Histocompatibility Complex (MHC) and show that, in general, both the MHC and immune-associated loci are significantly enriched for evidence of selection. Finally, we show that higher MHC diversity in farmers compared to hunter-gatherers is proportional to genome-wide diversity, which argues against selection for increased diversity.

This project demonstrates the power of aDNA for learning about selection, but also reveals its limitation. In particular, it is difficult to determine what specific factors are driving the selective events that we observe. Linking human and pathogen aDNA with archaeological and isotopic data provides a promising path for future work.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Infection and immune systems

OW-II3

Rapidly-evolving innate immune genes are younger, transcriptionally noisier and more pathologically dysregulated

Tzachi Hagai [1,*], Xi Chen [2], Ricardo Miragaia [2], Raghd Rostom [1], Tomas Gomes [2], Natalia Kunowska [2], Sarah Teichmann [2]

[1]EMBL - Europen Bioinformatic Institute, [2]Wellcome Trust Sanger Institute, Cambridge, United Kingdom

Abstract: The immune system is under constant pressure to evolve in the face of rapidly changing pathogens. At the same time, imbalanced changes in the immune response can lead to autoimmune disease. How these conflicting demands have shaped human immunity is not well understood.

Here, we compare the innate immune response of dermal fibroblasts from primates and rodents to two stimuli signalling the presence of viruses – double-stranded RNA and Interferon. Using single cell transcriptomics combined with chromatin activity analysis, we characterize the set of genes that rapidly diverged in expression between species. We show that genes that are related to inflammation and cellular defence diverged faster than genes with regulatory functions, and that faster divergences in expression between species correlates with higher transcriptional heterogeneity among individual cells. Expression divergence is reflected in conservation of chromatin sequence and activity. Furthermore, genes that diverged rapidly in expression experienced faster evolution in their coding sequence, are younger and had faster rates of gene gain and loss in the course of mammal evolution. Finally, rapidly evolving genes have stronger association with pathologies linked to dysregulation of this pathway.

Our study reveals that a group of immune genes with specific functions evolved rapidly in independent evolutionary modes, is expressed more stochastically and is associated with related immune diseases.

Expanded summary*: My research interests focus on how host and virus co-evolve.

In my previous work I studied viral mechanisms of molecular mimicry, where viruses take on some of their host's features in order to successfully avoid host surveillance and to rapidly replicate within the host cells. In that work we studied how viruses evolved to take advantage of some of the weakest points of our immune system.

This led me to my current work, where I seek to understand how continuous infections by diverse pathogens have shaped our immune system and how this affects our susceptibility to immune diseases.

The immune system is thought to be one of the fastest evolving systems due to selective pressure from rapidly evolving pathogens. In various genomic scans immune genes have been shown to have signatures of positive selection in their coding sequences, suggesting that immune genes evolve rapidly in their protein sequence as part of their evolutionary arms race with pathogens. However, most of the changes between species are not at the coding sequence level, and differences in other evolutionary modes, such as gene expression, might also contribute to functional differences between species immunity.

In this project, we set out to characterize how cell-intrinsic immunity – the expression programme that is upregulated in response to invading viruses – has changed in different evolutionary modes between closely related mammals and across the mammalian clade. For this, we have profiled the response of dermal fibroblast from primates and rodents to viral signals, using population and single-cell transcriptomics, giving us the ability to characterize temporal changes in gene expression at the single-cell level in each of these species. In parallel, we characterized this response at the chromatin level, by comparing histone markers activity between the different species.

This combined approach enabled us to characterize the set of genes that rapidly diverged in expression between species, showing that these genes are enriched with specific functions such as defence and inflammation. Furthermore, we observed that genes that diverged rapidly in expression tend to have higher stochasticity in their expression among individual cells in the population.

By comparing the divergence in expression to other evolutionary modes, we observe that genes that rapidly diverged in expression experienced faster coding sequence evolution. Moreover, these genes tend to have higher rates of gene duplication and loss across the mammalian clade, and as a result, these genes are younger.

Thus, we observe that a specific set of genes that is upregulated in response to viral signals diverged faster between species, is evolutionary younger and has higher variability in expression among responding cells. The fact that higher divergence is linked with

younger evolutionary age and with noisier transcription might point to spurious expression with no important functional outcome. However, the fact that genes that have diverged rapidly in their expression are also those involved in cellular defence and inflammation might suggest that the changes we observe are driven by selection and might have important roles in the ability to fend off pathogens.

Changes in both sequence and expression may lead to higher susceptibility to autoimmune diseases. We thus investigated how genes that are related to abnormal upregulation of the antiviral pathway evolved. Surprisingly, we observed that genes that are linked to misregulation of the pathway rapidly evolved in both sequence and expression. These results might suggest that some of the changes driven by pathogens have led to higher susceptibility to autoimmune disease.

Our work gives insights into the different modes in which the immune response evolved and how these changes might be linked to autoimmune disease. In the future, I plan to expand on this preliminary system, studying additional species and using other physiological systems.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-113

Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America

Kelly Swarts [1 2], Rafal Gutaker [1], Verena Schuenemann [3 4], Bruce Benz [5], Michael Blake [6], Robert Bukowski [7], James Holland [8 9], Melissa Kruse-Peeples [10], Nick Lepak [11], R G Matson [6], Lynda Prim [10], Cinta Romay [11], Jeffrey Ross-Ibarra [12], Jose De Jesus Sanchez-Gonzalez [13], Chris Schmidt [10], Evan Sofro [10], Johannes Krause [3 14], Detlef Weigel [15], Edward Buckler [2 16], Hernán Burbano [1,*]

[1]Research Group for Ancient Genomics and Evolution, Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tuebingen, Germany, [2]Department of Plant Breeding and Genetics, Cornell University, Ithaca, United States, [3]Institute of Archaeological Sciences, [4]Senckenberg Center for Human Evolution and Paleoenvironment, University of Tuebingen, Tuebingen, Germany, [5]Department of Biology, Texas Wesleyan University, Forth Worth, United States, [6]Department of Anthropology, University of British Columbia, Vancouver, Canada, [7]Bioinformatics Facility, Institute of Biotechnology, Cornell University, Ithaca, [8]Department of Crop and Soil Sciences, North Carolina State University, [9]USDA-ARS, Raleigh, [10]Native Seeds/SEARCH, Tucson, [11]Genomic Diversity Facility, Institute of Biotechnology, Cornell University, Ithaca, [12]Department of Plant Sciences, University of California, Davis, United States, [13]Centro Universitario de Ciencias Biológicas y Agropecuarias, Universidad de Guadalajara, Zapopan, Mexico, [14]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, [15]Department of Molecular Biology, Max Planck Institute for the Science of Human History, Tuebingen, Germany, [16]USDA-ARS, Ithaca, United States

Abstract: People introduced maize (Zea mays ssp. mays) to the southwestern US by 4,000 years ago. Full maize agriculture was established quickly in the lowland deserts but delayed in the temperate uplands for 2,000 years. Here, we hypothesized that this delay was caused by the necessity to select for early flowering, a characteristic of agriculturally important modern temperate maize. We sequenced fifteen 1,900-year old maize cobs from Turkey Pen Shelter in the temperate Southwest (contemporary Utah, USA). Genomic prediction models trained on diverse inbred lines and validated in modern landraces predicted that Turkey Pen maize was early flowering and therefore marginally adapted to its local environment. Population genetic analyses suggested temperate adaptation drove modern population differentiation and adaptive alleles were selected in situ from ancient standing variation. We showed that validated prediction of polygenic traits in crops improves our understanding of ancient phenotypes and opens up new avenue towards understanding our history and those of the animals and plants we domesticated.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA5

Functional implications of Neandertal introgression in modern humans

Michael Dannemann [1,*], Janet Kelso [1], Kay Prüfer [1]

[1]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Abstract: Between 1% and 6% of the genomes of present-day non-Africans derive from Neanderthals and Denisovans. A number of studies have provided evidence that this introgression from archaic humans contributes to important human phenotypes such as altitude adaptation, skin and hair physiology and immunity. In a number of these cases, the introgressed alleles show regulatory activity and are associated with expression changes of nearby genes, rather than with changes in protein sequences. We have explored the extent to which introgressed alleles contribute to differences in gene expression in multiple tissues between present-day humans, and compare these results to introgressed alleles that change the protein sequence. Finally, we link the potentially functional introgressed alleles that we identify to particular phenotypes via genome-wide association studies in humans and identify a number of introgressed alleles that have consequences for immune, neurological and metabolic phenotypes in present-day humans.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-123

Ancient DNA from Two Pre-Columbian Mummies from Sierra Tarahumara

Viridiana Villa-Islas [1,*], Cristina Valdiosera [2], Rosa Fregel [3], Alexandra Sockell [3], Mattias Jakobsson [4], Andres Moreno-Estrada [5], Carlos Bustamante [3], María C Ávila-Arcos [1]

[1]Population and Evolutionary Genomics Lab, International Laboratory for Human Genome Research, UNAM, Mexico, Queretaro, Mexico, [2]Department of Archaeology and History, La Trobe University, Melbourne, Australia, [3]Department of Genetics, Stanford University, California, United States, [4]Department of Evolutionary Biology, Uppsala University, Uppsala, Sweden, [5]Human Evolutionary and Population Genomics Lab, Laboratorio Nacional de Genómica para la Biodiversidad (LANGEBIO),, Irapuato, Mexico

Abstract: The Tarahumara are an indigenous population also known as Rarámuri, who inhabits the Sierra Tarahumara, mainly in the state of Chihuahua, Mexico. This millenary group is recognized for their incredible physical endurance and ability to run long distances. Genomic studies of this population and its ancestors may give us insights into their population history. Here we report the retrieval and sequencing of aDNA from two ca. 900 year-old pre-Columbian mummies found in a cave from Sierra Tarahumara. We performed Whole-Genome-Capture on the aDNA libraries to enrich the amount of endogenous DNA and sequenced until we reached saturation. We obtained enough data to cover two-thirds of the genome for one of the mummies and less than 10% for the second. This allowed us to perform PCA to compare the ancient individuals with modern Native Mexican groups. Preliminary analyses show that the mummies cluster with different present-day populations, respectively. This opens up questions of historical and anthropological interest; specifically regarding past genetic structure and migration between different indigenous groups.

Expanded summary*: The Tarahumara are an indigenous population also known as Rarámuri, who inhabits the Sierra Tarahumara, mainly in the state of Chihuahua, Mexico. Long-distance running is a cultural practice that men, women and children have practiced for centuries through the rugged landscape of the Sierra Tarahumara. Motivated by the interest in investigating a possible genetic basis for this extraordinary capacity, and to learn more about their population history, we launched a genomic study of this population combining genomic information from past and present individuals.

Ancient DNA (aDNA) offers an unparalleled source of information to better understand the evolutionary processes that have generated the genetic diversity of today's populations and to detect possible genes that could have been subject to selection in "real time". To gain insights into the population history of the Rarámuri and to explore possible signatures of adaptive evolution, we increased the sequencing depth of two ancient genomes belonging to two ca. 900-year old pre-Columbian mummies initially reported in Raghavan *et al*, 2015. Both mummies were found in a cave from Sierra Tarahumara; their DNA was extracted, built into Illumina libraries and sequenced at low depth. We increased the coverage of their genomes by generating additional libraries and performing Whole-Genome Capture (WGC) to enrich their endogenous content.

We obtained enough data to cover two-thirds of the genome for one of the mummies and less than 10% for the second. This allowed us to determine the mitochondrial haplogroup of the two individuals as C, and C1c1a, respectively; both are typical mitochondrial haplogroups of Native Americans. Also, we performed PCA to compare the ancient individuals with a reference panel of modern Native Mexican groups. Interestingly, these preliminary analyses show that one mummy clusters closely with present-day Tarahumaras, while the second clusters with a geographically distant population. This result opens up questions of historical and anthropological interest; specifically regarding past genetic structure and migration between different indigenous groups and demands further and more detailed analyses of the genomic data in a population context.

In the next phase of the project we will increase the depth of coverage of the genomes and combine with knowledge generated from modern Rarámuri. We are in parallel characterizing functional variation in present-day Rarámuri and identifying targets of adaptive

evolution. The combination of both sources of genetic information—ancient and modern—might help characterize the temporality of the variants associated with adaptive evolution, which might, or might not, be related to physical endurance.

This study, not only complements our knowledge on the genetic component of the Tarahumara population, but also contributes to a better understanding of the pre-Columbian Mexican native populations, which have been little studied from the point of view of genetic diversity. Few studies to date have focused on the paleogenomic study of ancient human samples in the Americas and none has characterized complete ancient genomes of samples from Mexico despite its rich historical and cultural heritage as reflected in the vast archaeological record. Consequently, paleogenomic studies in Mexico have a great potential.

Furthermore, another important contribution of this study is the implementation of the WGC enrichment method, which has been proposed to increase the endogenous DNA content of ancient human samples, yet it has not been tested thoroughly. This work allows testing new parameters of the protocol and gain further insights about its performance on mummified samples.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-121

A pre-existing isolation by distance gradient in West Eurasia may partly account for the observed “steppe” component in Europe

Luca Pagani [1,*], Lehti Saag [1], Anto Aasa [2], Flora Jay [3], Mait Metspalu [1]

[1]Estonian Biocentre, [2]Department of Geography, University of Tartu, Tartu, Estonia, [3]LRI, Paris-Sud University, Paris, France

Abstract: It has been proposed that modern European populations can be modelled, by and large, as a three-way mixture of Hunter-Gatherer, Anatolian Neolithic and Steppe components that took place after 6kya (Haak et al. 2015, Allentoft et al. 2015). Particularly the pre-existing Hunter-Gatherer are thought to have admixed with incoming Early Neolithic people from Anatolian and, subsequently, with people carrying a “Steppe” component from the East. These people were likely bearing the so called Yamnaya and/or Corded-Ware cultures, and their initial impact of the European gene pool was estimated to be as high as 75% (Haak et al. 2015).
However ancient DNA samples from East European and Caucasian Hunter-Gatherers as well as from Early Iranian Neolithic, dating from before the Yamnaya expansion, already show signs of this so called “Steppe” component (Lazaridis et al. 2016). Such an observation is compatible with the presence of a pre-existing genetic gradient ranging from Caucasus/Iran all the way to Europe, which likely formed through isolation by distance over thousands of years.
Here we show that such a gradient, defined as decrease of “steppe” component with distance from Iran, can be inferred from ancient samples pre-dating the Yamnaya expansion ($r^2=0.93$).
When analysed in the light of this gradient, later ancient and modern samples from Europe still display an excess of Steppe component, however this excess is less pronounced than previously estimated. Additionally we found that, of the analysed samples, modern South Asians show the highest excess of “steppe” component, pointing to the documented, recent links between the Caucasus/Iran populations and the South Asian peninsula.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-118

Population genomics of wild yeast over decades of evolution in southeastern Pennsylvania

Joshua A Shapiro [1,*]

[1]Bryn Mawr College, Bryn Mawr, United States

Abstract: The genomic diversity of *Saccharomyces cerevisiae* has been studied extensively in a global context, through comparisons of isolates from a variety of natural, industrial, and clinical settings. Such studies have tended to find limited evidence for positive selection across the species. By contrast, studies of experimental evolution in laboratory environments have found the species to adapt rapidly, both through changes at individual nucleotides and through larger structural changes such as segmental duplications and aneuploidies. To connect the patterns of evolution observed at these divergent scales, we took advantage of historical isolations of *S. cerevisiae* from oak trees collected nearly twenty years ago in southeastern Pennsylvania, comparing those isolates to recent collections from the same region and environment. Given the expected generation time for *S.* cerevisiae, these two populations are likely separated by thousands of generations of evolution. Through genomic sequencing, using both short and long read technologies, we identify and characterize nucleotide and structural variants that have changed in frequency between these time-separated populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-124

Complete mitochondrial genomes provide additional evidence on the geographical origin of the indigenous people of the Canary Islands

Rosa Fregel [1,*], Alejandra C. Ordóñez [2], André E. R. Soares [3], Jonathan Santana [4], Matilde Arnay [2], Beth Shapiro [3], Carlos D. Bustamante [1]

[1]Department of Genetics, Stanford University, Stanford, United States, [2]Department of Prehistory, Universidad de La Laguna, La Laguna, Spain, [3]Paleogenomics Lab, University of California Santa Cruz, Santa Cruz, United States, [4]Department of Prehistory, Universidad de Las Palmas de Gran Canaria, Las Palmas, Spain

Abstract: Deciphering the geographic origin of the Canary Islands' indigenous inhabitants has fascinated both scholars and the general public. Ancient DNA (aDNA) evidence, based on PCR techniques, has confirmed the presence of North African mitochondrial DNA (mtDNA) lineages in the indigenous people, including the North African U6 haplogroup. In fact, one striking result was the discovery of the U6b1a sub-haplogroup, which is exclusively observed in ancient and modern populations of the Canary Islands, and it is absent in North Africa. Classical aDNA techniques have provided valuable information, but results have been always hindered by the risk of modern contamination. Moreover, PCR-based analyses are limited to a small portion of the mtDNA genome and important information from the coding region is unavailable.
In this study, we apply for the first time next-generation sequencing to the recovery of whole mtDNA genomes of indigenous people of the Canary Islands (n=44). Most of the lineages observed in the ancient population of the Canary Islands belong to West Eurasian and North African haplogroups, confirming previous results. As expected from archaeological, anthropological and linguistic studies, the majority of indigenous mtDNA lineages are present in the Maghreb. Phylogenetic analysis indicates the presence of additional autochthonous lineages that mimic the distribution observed for U6b1a. Coalescence ages for those Canarian-autochthonous subhaplogroups are mostly in agreement with the colonization time proposed by radiocarbon dates and archaeological criteria. However, an older autochthonous lineage as U6b1a is unlikely to have developed in the Canary Islands based on currently available archaeological records.

Expanded summary*: The goal of this project is to apply paleogenomic techniques to the study of the Canary Islands' prehistory for the first time. During the 13th - 14th centuries, European sailors eagerly traveled the oceans searching for new worlds. The subsequent expansion of European colonies across the world, triggered the European dominance of the global economy, but also had important cultural and ecological consequences, because it brought together, for the first time, distant civilizations and environments. Portuguese sailors discovered several groups of islands in the Atlantic Ocean in the 13th century. Around this time, the Portuguese and Castilians began to settle the Atlantic archipelagos, including the Azores, Madeira and Cape Verde, but only the Canary Islands were found to be inhabited by an indigenous population, generally known as Guanches. During the 15th century, the Canary Islands were gradually conquered, directly or indirectly, by the Spanish kingdom of Castile, beginning with the island of Lanzarote in 1402 and finishing with Tenerife in 1496. In general terms, the Conquest was exceptionally violent, due in part to the fierce resistance of the indigenous people against the invaders. The crushing of the resistance, and the subsequent European colonization, had a great impact on the indigenous way of life. In spite of the indigenous protective policy of Queen Isabel 'La Católica', who legally abolished slavery on the Islands in 1498, a large number of Guanches were deported during and after the Conquest, and some of them were introduced into the 16th century European slave trade. Those that survived and stayed within the islands progressively mixed with the European colonizers, leading to the loss of indigenous culture and language.
Most archaeological, anthropological and linguistic researchers point to a North African origin for the Canary indigenous people, more precisely related to the proto-Berber and Berber world. Ancient DNA analyses on the Guanche population using classical PCR-based methods have confirmed the presence of North African lineages in the indigenous people, including different sublineages of the characteristic North African U6 haplogroup. One important result was the characterization of the U6b1a sub-haplogroup, which is exclusively observed in ancient and modern populations of the Canary Islands, and not in North Africa. However, due to the lack of

samples from the eastern islands (Gran Canaria, Lanzarote and Fuerteventura) and to limitations associated with the use of only a small portion of the mtDNA genome, we were unable to identify a specific geographical origin for the Guanche people.

In this project, we used next-generation sequencing to generate complete mtDNA genomes from the indigenous population for the first time. We obtained high-coverage mtDNA genomes for 44 human remains excavated from 23 different archaeological sites distributed across the entire Canarian archipelago. Most of the lineages observed in the ancient population of the Canary Islands belong to West Eurasian (H, J and T) and North African (U6) haplogroups, confirming previous results using PCR techniques. As expected from archaeological, anthropological and linguistic studies, our results indicate that the first inhabitants of the Canary Islands are related to modern populations of North Africa. However, the absence or low frequency of some key haplogroups in North Africa indicates that the continental genetic composition has been modified by later human migrations. More strikingly, phylogenetic analysis indicates the presence of additional autochthonous lineages that mimic the distribution observed for U6b1a. By using whole-genome sequences of indigenous samples, we have been able to identify additional autochthonous lineages that mimic the distribution observed for U6b1a. Those Canarian-specific haplogroups are sublineages of the Eurasian H, J and T, and the African L3 macrohaplogroups. With this refined phylogeographic information we will be able to unequivocally assess the indigenous origin of maternal lineages observed in the modern Canarian population.

Apart from its clear significance for understanding the demographic history of the Guanche population, the results obtained in this project are also of paramount importance for the Canarian society. The success of projects and companies providing ancestry information indicates how people crave knowledge about their origin. This is especially true for the Canary Islands. The indigenous population plays an important role on the identity of the modern inhabitants of the Canary Islands, who are very interested in knowing as much as possible about this part of their history. However, due to the mystical aura that has always surrounded the Guanche population since the first European chroniclers started writing about them, misleading and pseudo-scientific information is sometimes fed to the public and accepted as fact. It is the responsibility of scientists to provide society with evidence and help providing insight to differentiate what is fact and what is myth. This project will allow us to keep answering those questions with state-of-the-art methods in the field.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA1

Modelling rare genetic variation in modern and ancient genomes to study recent human history.

Stephan Schiffels [1,*]

[1]Department for Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

Abstract: Recent studies have provided key insights into human history by analysing human genetic variation, including major migration events and population turnovers. However, for many historical questions, the events of interest involve populations that are very closely related with each other, which pushes many traditional methods to the limits of resolution. Here I will present new methodology that uses rare genetic variation to distinguish closely related ancestries. In particular, our new method rarecoal infers population history and identifies fine-scale genetic ancestry from rare variants, by probabilistically modelling the coalescent tree of rare derived mutations under a population model with split times, branch population sizes and admixture edges. I will present two applications: First, I present a joint analysis of 19 ancient English samples that were published in two studies in 2016 (Schiffels et al. 2016, Martiniano et al. 2016), ranging from the late Iron Age to the middle Anglo-Saxon period. When analyzed together with hundreds of modern European genomes, we gain detailed insights into population history and admixture events in England during the last 2,000 years. In the second study, we analyse hundreds of modern samples from Siberia and America, together with published and unpublished ancient genomes, and gain insights into the prehistory of Native Americans. In particular, we show evidence based on rarecoal and other methods, that Athabaskan speaking Native Americans have ancestry from three independent migration events into America, including Palaeo-Eskimos and Inuit.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-110

7,000 years of change: Migration and admixture in the population history of the Caribbean

Maria A Nieves-Colon [1,*], William J Pestle [2], Jada Benn-Torres [3], Carlos D Bustamante [4], Anne C Stone [1]

[1]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, [2]Department of Anthropology, University of Miami, Coral Gables, FL, [3]Vanderbilt University, Nashville, TN, [4]Department of Genetics, Stanford University, Stanford, CA, United States

Abstract:

Although the Caribbean has been continuously inhabited for the last 7,000 years, European contact in the last 500 years dramatically reshaped the cultural and genetic makeup of island populations. Several recent studies have explored the genetic diversity of Caribbean Latinos, and have characterized Native American variation present within their genomes. However, the difficulty of obtaining ancient DNA from pre-contact populations and the underrepresentation of non-Latino Caribbean islanders in genetic research, have prevented a complete understanding of genetic variation over time and space in the Caribbean basin. Here we discuss research that takes two approaches towards characterizing migration and admixture in Caribbean populations: an ancient DNA analysis of 139 individuals from three pre-contact archaeological sites in Puerto Rico (A.D. 500–1300), and an analysis of whole genome variants from 55 Afro-Caribbeans in five Lesser Antillean populations. Our ancient DNA analysis traces the origin and number of pre-contact migrations to Puerto Rico and examines the extent of genetic continuity between ancient and modern populations. In contrast, our modern DNA work analyzes autosomal SNP genotypes to characterize complex patterns of admixture since European contact among Lesser Antillean Afro-Caribbeans. Our findings characterize how ancient indigenous groups, European colonial regimes, the African Slave Trade and modern labor movements have shaped the genomic diversity of Caribbean islanders. In addition to its anthropological or historical importance, such knowledge is also essential for informing the identification of medically relevant genetic variation in these populations.

Expanded summary*:

Characterizing how migration and admixture shapes human genetic diversity is vital for understanding human evolution, history and health. This is especially true in world regions that have undergone recent and dramatic demographic shifts, such as the Caribbean. Previous research with admixed Caribbean populations has shown that many islanders retain genomic variation from pre-Columbian indigenous groups, but also carry signatures of more recent admixture events fostered by European colonization and the African Slave Trade. However, a complete understanding of human genomic diversity across the Caribbean region is hampered by sampling gaps of both past and present populations. Due to the difficulties of obtaining ancient DNA (aDNA) from the tropics, the genetic diversity of pre-Columbian Caribbean groups is not well characterized. Efforts have been made to address this problem by studying Native American fragments in the genomes of admixed islanders. But, because modern populations do not retain all the genomic diversity of ancient groups, this approach provides limited resolution for reconstructing ancient demographic events. Further, many Caribbean populations remain underrepresented in large catalogs of genomic variation. Except for Barbadian Afro-Caribbeans, recently included in 1000 Genomes Phase 3, genetics research on Lesser Antillean populations has been limited to uni-parental loci and low-density ancestry informative markers. The present research seeks to fill in these gaps through two approaches: an aDNA analysis of 139 individuals from three archaeological sites in Puerto Rico (A.D. 500–1300), and an analysis of genome-wide SNP variants from 55 Afro-Caribbeans in five Lesser Antillean (LA) populations.

The aDNA investigation characterizes patterns of migration and genetic admixture in pre-Columbian Puerto Rico, and examines the extent of genetic continuity between ancient groups and modern islanders. In-solution capture and next-generation sequencing were used to obtain ancient DNA from 139 human skeletal remains (dated between A.D. 500–1300), from the sites of Tibes (n=52), Paso del Indio (n=50) and Punta Candelero (n=37). Preliminary data obtained from 24 complete mitochondrial genomes (mean read depth: 9.8x) suggest that pre-Columbian communities in Puerto Rico share genetic affinity with several extant South American and Mesoamerican indigenous populations. We also find that most pre-Columbian mtDNA lineages are not present in the Americas today, except for one, which is found almost exclusively in modern Puerto Ricans. These data support an origins scenario of complex and

continuous admixture for ancient Caribbean groups but also underscore the large effect that contact-era population declines had on indigenous communities. Autosomal genotypes currently being generated from these remains will further inform these issues.

The second part of our project analyzes autosomal SNP genotypes in 55 self-identified Afro-Caribbeans from St. Kitts (n=5), St. Lucia (n=10), St. Vincent (n=15), Grenada (n=6), and Trinidad (n=19). We characterize patterns of genome-wide variation and ancestry in these individuals and compare them to exising data from other recently admixed American populations. We observe a complex pattern of admixture among in the LA Afro-Caribbeans with inputs from up to five continental sources and strong signatures of sex-biased mating. African ancestry proportions are high, but Native American ancestry is extremely low. This pattern contrasts sharply with that observed in Caribbean Latinos and is more similar to that observed in Haitians and Barbadians. We further observe that Trinidadian Afro-Caribbeans have the highest proportion of admixture with East and South Asian populations of all Caribbean populations studied to date.

Overall, our findings underscore the large impact of post-contact demographic shifts on Caribbean population history and illustrate how genomic diversity has changed in this region over the last 7,000 years. In addition, this work increases the representation of admixed and diverse populations in available genomic datasets and has the potential to inform future functional and clinical genetics research with admixed Caribbean islanders.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-111

Revealing the impact of the Younger Dryas on small vertebrate diversity of the Edwards Plateau, Texas, through aDNA bulk bone metabarcoding

Daniel Werndly [1,*], Thomas Stafford [2], Ernest Lundelius [3], Alicia Grealy [1], Erin Keenan Early [4], Michael Bunce [1]

[1]Environment and Agriculture, Curtin University, Perth, Australia, [2]Stafford Research, Lafayette, [3]Jackson School of Geosciences, University of Texas, Austin, [4]Gault School of Archaeological Research, Texas State University, San Marcos, United States

Abstract: The extinction of North American megafauna took place in the 2 ka. preceding the Younger Dryas, a period of cooling circa. 11.7 ka. to 8 ka. The effect this cooling period had on smaller vertebrates is poorly understood, as osteological remains of such vertebrates are usually cryptic or fragmentary, making taxonomic assignment through morphological assessment unreliable. Ancient DNA sequencing offers an alternate method for assigning species identifications from paleontological contexts, and is used to discern morphologically cryptic species and identify multiple taxa within mixed samples simultaneously. From excavations in Hall's Cave, a limestone cave in the centre of the Edward's Plateau in Texas, 30 bulk samples of 100 bones were collected representing the small vertebrate diversity spanning the Last Glacial Maximum ca. 18 ka. to the Middle Holocene ca. 6 ka. Sequence data was recovered from extinct or extirpated mammals, squamates and aves as well as extant vertebrate species through the bulk bone metabarcoding technique. This was used to explore the paleobiotic changes between the Pleistocene and the Holocene. These data were subsequently tested to ascertain if changes occurred rapidly, through a drop in diversity during the Younger Dryas, or gradually, through a reduction in diversity in the 5 ka. prior to the Younger Dryas. Preserved ancient DNA is rare in central USA and our data are crucial for a comprehensive reconstruction of the Edward's Plateau paleoenvironment. Notably, these data facilitate the modeling of species' responses to climatic change, which is vital for understanding extinction processes and present-day conservation biology.

Statement: I commenced my Honours year in August of 2016 with the aim of using ancient DNA to generate a paleoenvironmental record of central Texas. I spent the first month preparing for an extended field trip to excavate fossil bone in September and October. After the two weeks spent at Hall's Cave, a formation in the centre of the Edward's Plateau, working with a number of researchers based in Texas and the wider U.S. I returned to Austin to continue sieving the excavated material and pick out bones and any artefacts. During this time I was working in the University of Texas, Austin, Vertebrate Paleontology Laboratory. I have since returned to Curtin University in Perth to start the biomolecular element of my thesis and generate sequence data. On completion of my thesis in June, I hope to show that ancient DNA can be a valuable tool for isolating and understanding mechanisms for extinction and extirpation events.

By returning to Austin for the conference, I will be able to reconnect with the researchers and academics with whom I took residence while in Texas, and potentially progress the my research of the Younger Dryas extinction event even further by collecting more material from more diverse sites throughout Texas. I will also experience a professional conference environment, and gain access to a broad spectrum of cutting edge research in the fields of molecular biology and evolution. This will surely influence my research for the better.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-140

Ancient HLA immune gene polymorphism in pre-European contact Native American populations

Federica Pierini [1,*], Austin W. Reynolds [2], Jaime Mata-Míguez [2], Lisa Böhme [3], Marcel Nutsua [3], Almut Nebel [3], Ben Krause-Kyora [3 4], Deborah A. Bolnick [2], Tobias L. Lenz [1]

[1]Max Planck Institute for Evolutionary Biology, Plön, Germany, [2]University of Texas at Austin, Austin, United States, [3]Institute of Clinical Molecular Biology, Kiel University, Kiel, [4]Max Planck Institute for the Science of Human History, Jena, Germany

Abstract: It has been proposed that selection for resistance to infection drives the evolution of genetic variability at the major histocompatibility complex (MHC). In humans, a number of studies suggest that specific alleles of the human leukocyte antigen (HLA) system are associated with susceptibility or resistance to infectious diseases (i.e. HIV, malaria, tuberculosis, hepatitis) but, unlike for several other species, convincing evidence for pathogen-driven selection is still awaited. In this light, the recent development of genomic tools for analysis of ancient DNA (aDNA) provides a unique opportunity to unravel selection processes throughout human history.

Archeological, historical and genetic studies indicate that Native Americans experienced a strong population bottleneck following European contact. It has been proposed that Native Americans' HLA genes may have lacked both genetic polymorphism generally and specific resistance alleles to a variety of new pathogens introduced by European colonizers, resulting in an increased susceptibility to new diseases (i.e. smallpox, measles, diphtheria, rubella, and mumps).

Here we use in-solution targeted capture on aDNA samples to analyze HLA polymorphism in an ancient Native American population. Preliminary analyses reveal allelic lineages currently present in Amerindian and/or Asian populations. Using this data we investigate potential HLA allele frequency shifts from pre- to post-European contact. This work is expected to highlight possible signatures of selection in response to new pathogens introduced by European colonizers.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-141

Genome wide data from the Iron Age provides insights into the population history of Finland.

Thiseas Christos Lamnidis [1,*], Kerttu Majander [2], Elina Salmela [1 3], Anna Wessman [4], Antti Sajantila [5], Päivi Onkamo [3], Stephan Schiffels [1], Johannes Krause [1 2]

[1]Department of Archaogenetics, Max Planck Institute for the Science of Human History, Jena, [2]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany, [3]Department of Biosciences, [4]Department of Archaeology, [5]Department of Forensic Medicine, University of Helsinki, Helsinki, Finland

Abstract: The population history of Finland is subject of an ongoing debate, in particular with respect to the relationship and origins of modern Finnish and Saami people. Here we analyse genome-wide data, extracted from three teeth found in the archaeological site of Levänluhta, in southern Ostrobothnia. The site dates back to the Iron Age between 550-800 AD, according to the artefacts recovered, while radiocarbon dating on scattered femurs from the site span 350-730 AD. When analysed together with previously published ancient European samples and with modern European populations, the ancient Finnish samples lack a genetic component found in early Neolithic Farmers and all modern European populations today. Instead, we find that they are more closely related to modern Siberian and East Asian populations than modern Finnish are, a pattern also observed in genetic data from modern Saami. Our results suggest that the ancestral Saami population 1500 years ago, inhabited a larger region than today, extending as far south as Levänluhta. Such a scenario is also supported by linguistic evidence suggesting most of Finland to have been speaking Saami languages before 1000 AD. We also observe genetic differences between modern Saami and our ancient samples, which are likely to have arisen due to admixture with Finnish people during the last 1500 years.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA8

Targeted re-sequencing of historic and modern samples reveals the complexity of potato introduction to Europe

Rafal Marek Gutaker [1,*], José Luis Fernández Alonso [2], Salomé Prat [3], Sandra Knapp [4], Hernán Andrés Burbano [1]

[1]Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tuebingen, Germany, [2]Departamento de Biodiversidad y Conservación, Real Jardín Botánico-CSIC, [3]Departamento de Genetica Molecular de Plantas, Centro Nacional de Biotecnologia-CSIC, Madrid, Spain, [4]Department of Life Sciences, The Natural History Museum, London, United Kingdom

Abstract: Potato was domesticated in the South American Andes and was initially introduced to Europe in the 16th century. Over time it played increasingly important role as a food source. To tuberize before winter in Europe, potato had to adapt to shorter growing season through photoperiod insensitivity. Molecular mechanism of such adaptation in modern cultivars is thought to be exclusively attributed to structural variation in *StCDF* gene. We collected herbarium specimens of European cultivated potato dated between 1660-1899 AD to investigate the genomics of potato introduction in Europe. We successfully optimized extraction protocols to facilitate library-based re-sequencing of highly degraded DNA from old herbarium samples. Subsequently, we utilized hybridization array to capture regions with curated nuclear SNPs, whole chloroplast genome and over 300 genes related to photoperiod response. Comparison of nuclear SNPs and chloroplast assemblies between historic and modern potato samples unveiled waves of migrations that changed European gene pool between 19th and 20th century. We uncovered increase in genetic diversity of European potato in last hundred years, suggesting continuous influx of South American germplasm. We showed that historic European samples carry a structural variant of *StCDF* that confers adaptation to European climate, however, only at low frequencies. Finally, we report changes in nucleotide diversity in other genes that could be relevant to the process of adaptation. In summary, we present an interesting case of crop's recent adaptation to new environment and provide compelling evidence for the complexity of such scenarios.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-127

Ancient DNA damage mixture patterns as an estimator of autosomal contamination rate

Philip Johnson [1,*]

[1]Department of Biology, University of Maryland, College Park, United States

Abstract: Contamination by modern humans remains a nagging concern for ancient DNA studies, particularly when studying ancient hominins. Ancient DNA damage in the form of C->T base lesions provides a rough signal of authenticity, but the dream of using damage to directly estimate contamination has proved elusive due to the variability in true damage rates between ancient samples. Existing approaches for contamination detection use mitochondrial DNA, sex chromosome ratios, differential allele frequencies, and excess levels of polymorphism indicative of ploidy > 2. However, these methods all have flaws when the target of interest is the autosomal nuclear genome. For instance: the mitochondrial:nuclear genome ratio varies wildly between cell types; the power of sex chromosomes depends on the endogenous sex and contaminator sex; allele frequency differences virtually disappear when the "ancient" sample is relatively recent; and low but still troublesome levels of contamination leave very little signal of excess polymorphism. Here, we present a new method that explicitly uses ancient DNA damage patterns in the sequence reads to identify a mixture signal in which some proportion of the data is damaged (i.e., authentic) and some proportion is not (i.e., contamination).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA9

Functional characterization of Neanderthal introgression

Natalie Telis [1,*], Kelley Harris, Jonathan Pritchard [2 3 4]

[1]Biomedical Informatics, Stanford, Stanford, [2]Genetics, [3]Biology, Stanford, Stanford, CA, [4]Howard Hughes Medical Institute, HHMI, Chevy Chase, MD, United States

Abstract: There has been a long history of interest in the extent of, and selective consequences of, human interbreeding with other hominins. Existing knowledge of Neanderthal introgressed variants across the genome has revealed several deserts, as well as a general depletion in the vicinity of genes, with weak association with testes expression. Characterizing the functional impacts of Neanderthal introgression can shed light into broader functional trends of modern noncoding selection. By examining distributions of introgression across regulatory elements, we identify genome-wide depression of Neanderthal allele frequencies in promoters and enhancers. This genome-wide depletion increases in extremity with increasing allele frequencies. Alongside this broad depletion of Neanderthal introgression in regulatory regions, we confirm independent tissue-specific signals associated with development. Based our our analysis, we present a model linking modern regulatory consequence with fitness and distribution of existing Neanderthal variations.

Expanded summary*: Introduction. Inferring coalescent times of genomic samples has been incredibly informative with regards to the understanding of human demographic history. However, it is clear that coalescent times may provide a great amount of information across loci. These local coalescent times may provide information about natural selection as well as historical demography.

Applications. Locus-specific coalescent times may be be used for two primary applications: (1) to examine locus-specific selection and (2) to infer detailed population demography. (1) With regards to natural selection, it is clear that locus-specific coalescent times will reflect locus-specific selection signals. For example, a long history of balancing selection at a locus would result in a much deeper time to coalescence than would be observed genome-wide. Likewise, a recent selective sweep of a particular allele at one locus would result in a significantly altered sequence tree with a much shallower coalescent time than that observed genome-wide. Although the latter can be detected through haplotype-based selection scans, the former cannot. (2) The distribution of locus-specific coalescent time allows us to infer demography of given population samples, allowing us to extract information about the relationships of modern population samples with ancient samples.

Method. Our method uses a simple estimator (Thomson et al. 2002; Hudson 2007) and leverages unphased genotyping data to calculate a window-wide estimate of the time to most recent common ancestor (TMRCA). From these window estimates we infer change points to identify blocks of windows with likely identical underlying trees. We use this approach on combinations of modern and ancient samples to examine the genome-wide distribution of locus-specific coalescent times, which enables us to (a) examine signals of selection and (b) understand the demography shared by modern and ancient hominid samples.

Significance. As our method is not dependent on whole-genome sequence but only requires genotyping, it can be used on samples with extremely poor sequencing quality, like may ancient DNA samples. This enables detailed examination of modern-ancient sample demography. Moreover, our method provides a novel way to detect selection, which improves on methods of detecting directional selection by additionally providing a way to detect balancing selection. Finally, since it is reliant on genotyping data and not sequencing data, it can be applied to an unprecedented number of samples to look at population-specific demography and selection.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-128

Simultaneous Estimates of Archaic Admixture and Ancient Population Sizes

Alan Rogers [1,*]

[1]Anthropology, University of Utah, Salt Lake City, United States

Abstract:

To estimate archaic admixture, one must control for the sizes and separation times of ancient populations. We describe a new method that provides simultaneous estimates of these parameters in complex models of population history. Preliminary results confirm several previous results, but indicate that (1) Papuans have more Neanderthal admixture and less Denisovan admixture than previously thought; and (2) the archaic populations that contributed genes to modern humans were much larger than previous estimates.
This work was supported by grant BCS-638840 from the National Science Foundation.

To estimate archaic admixture, one must control for the sizes andseparation times of ancient populations. We describe a new method thatprovides simultaneous estimates of these parameters in complex modelsof population history. Preliminary results confirm several previousresults, but indicate that (1) Papuans have more Neanderthal admixtureand less Denisovan admixture than previously thought; and (2) the archaicpopulations that contributed genes to modern humans were much largerthan previous estimates.
This work was supported by grant BCS-638840 from the NationalScience Foundation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POB-425

DEEP LEARNING FOR REFERENCE-FREE INFERENCE OF ARCHAIC LOCAL ANCESTRY

Arun Durvasula [1,*], Sriram Sankararaman [1 2]

[1]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, [2]Department of Computer Science, University of California, Los Angeles, United States

Poster: Statistical analyses of genomic data from diverse human populations have demonstrated that modern human populations trace a small proportion of their genetic ancestry to archaic hominins such as Neanderthals and Denisovans. These analyses were enabled by the availability of archaic genome sequences. Several studies have suggested that archaic admixture has been common in the history of populations even though the ancestral archaic populations have not been identified. This observation motivates the problem of reference-free archaic local ancestry inference, i.e., inferring segments of the genome that trace their ancestry to an archaic population even in the absence of reference archaic sequences.

Previous attempts at reference-free archaic local ancestry inference have relied on a limited number of features or summary statistics (such as S*), which have limited power and a high false positive rate. Recent advances in deep learning permit the learning of complex, non-linear features that can be useful in a number of inferential tasks.

Here, we present a deep neural network (DNN) for archaic local ancestry inference. The DNN learns a number of features from patterns of genetic variation across a number of human genomes that allows accurate inference of archaic inference, with an overall accuracy of 93% and an Area Under the Receiver Operator Curve (AUROC) of 0.98. The baseline AUROC for S* is 0.77. Preliminary analyses of a sub-Saharan African population find that an average of 2.03% (SD: 0.38) of their genomes is labelled as archaic, in line with previous estimates.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA7

Bridging the Divide Between Modern and Ancient DNA

David Reich*

Abstract: Genome-wide studies of human variation have for the most part focused either on DNA from present-day individuals, or from individuals who lived prior to 4,000 years ago. However, developing a detailed understanding of how the peoples who lived in the early Bronze Age contributed to Iron Age populations who in turn contributed to Medieval populations who in turn contributed to people living today, has been difficult. One challenge is that by the beginning of the Bronze Age (at least in Western Eurasia where the most ancient DNA data have been collected), the ancestry composition of many populations was very similar to that of populations that live in the same regions today. As a result, the powerful methods that have been developed for learning about population history based on allele frequency correlation patterns are sometimes not able to discern the often subtle differences in ancestry composition between past populations. In this talk, I will describe work in which my colleagues and I have tried to begin to bridge this divide, both by studying ancient samples from intermediate time points, and by deploying more sensitive statistical methods.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-122

The genetic history of the Indonesian Pygmies of Flores

Serena Tucci [1 2,*], Benjamin Vernot [3], Rajiv C. McCoy [1], Samuel Vohr [4], Matthew R. Robinson [5], Chiara Barbieri [6], Joshua Schraiber [7 8], Herawati Sudoyo [9 10], Peter M. Visscher [5 11], Guido Barbujani [2], Richard E. Green [4], Joshua M. Akey [1]

[1]Department of Genome Sciences, University of Washington, Seattle, United States, [2]Department of Life Sciences and Biotechnologies, University of Ferrara, Ferrara, Italy, [3]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [4]Department of Biomolecular Engineering, University of California, Santa Cruz, United States, [5]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia, [6]Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Jena, Germany, [7]Institute for Genomics and Evolutionary Medicine, [8]Department of Biology, Temple University, Temple, United States, [9]Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, [10]Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta, Indonesia, [11]Queensland Brain Institute, The University of Queensland, Brisbane, Australia

Abstract: Modern human pygmy populations are distributed globally, and their short stature is hypothesized to represent one aspect of a complex eco-geographic adaptation to rainforest or island environments. Although numerous genetic studies have been conducted on pygmies in Africa and Southeast Asia, to date, there have been no genome-scale analyses of the pygmy population living on the island of Flores, Indonesia. Intriguingly, this population lives in a village near the cave where remains of a small-bodied human species, *Homo floresiensis,* were recently found. Here, we describe whole-genome sequences (>40x) from 10 Flores pygmy individuals, as well as genome-wide SNP data from 35 individuals. The Flores genomes harbor on average 48 Mb and 4.4 Mb of Neandertal and Denisovan sequence, respectively. Height-associated loci identified in European populations are significantly differentiated in the Flores pygmies, who possess an excess of height-decreasing alleles and a deficiency of height-increasing alleles. This result is consistent with a hypothesis of polygenic selection acting on standing variation for reduced stature in Flores. Finally, we identify a strong signature of recent positive selection encompassing the FADS gene cluster on chromosome 11, encoding for fatty acid desaturases that regulate the metabolism of long-chain polyunsaturated fatty acids (LC-PUFA). Flores individuals are nearly fixed for an ancestral haplotype that is predicted to confer reduced capacity to synthesize LC-PUFA from plant-based precursors. Our results add to emerging evidence that the FADS region has been a recurrent target of selection in diverse human populations, possibly in response to changing diets.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-126

Integrating ancient DNA and large-scale clinical biobanks to study recent human evolution

Corinne Simonti [1], John A. Capra [2,*] and eMERGE Network

[1]Vanderbilt Genetics Institute, [2]Biological Sciences, Vanderbilt University, Nashville, United States

Abstract: Thousands of regions throughout the human genome have been identified in studies of positive selection, archaic admixture, and other evolutionary events that have shaped human populations. The increasing availability of ancient DNA samples is enabling resolution of the temporal and spatial dynamics of many of these evolutionary changes. However, the functions of the vast majority of these variants are unknown. Historically, determining the phenotypic effects of such variants relied on indirect evidence, such as the function of nearby genes, or low-throughput methods for functional characterization, such as enhancer reporter assays. To overcome this limitation, we have developed a high-throughput approach for characterizing the function of variation arising in recent human history that leverages clinical biobanks and electronic health records (EHRs). We mined the EHRs of ~28,000 genotyped individuals from seven hospitals across the US to identify cases and controls for a diverse set of ~1,500 clinical phenotypes.

To illustrate the use of this resource in evolutionary studies, we used ancient DNA data to predict the age and origin of hundreds of human-specific genetic variants. We focused on two classes of variant: those appearing on the human lineage that have risen to near fixation since divergence with chimp (hominin-derived) and variants that have appeared since the divergence of humans from our last common ancestor with Neanderthals (modern-human-specific). Using our biobank-linked EHR resource, we discovered and replicated associations of human-specific and hominin-derived variants with many evolutionarily relevant phenotypes, including skeletal disorders and risk of bone fracture. We also observe instances of archaic introgression reintroducing functional alleles lost in the out of Africa bottleneck. Using functional genomics data from ENCODE/Roadmap Epigenomics and gene expression data from GTEx, we demonstrate that these variants often influence the regulation of nearby genes relevant to the associated phenotypes. Our results establish that recent hominin-derived and human-specific variants influence risk for evolutionarily relevant phenotypes in modern humans, and demonstrate the utility of EHR data linked to genotypes to inform evolutionary analyses.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-115

Identification of large structural variants in archaic hominins

Laurits Skov [1,*], Mikkel Heide Schierup [1] and The Vindija Genome Analysis Consortium

[1]Bioinformatic research center, Aarhus university, Aarhus C, Denmark

Abstract: Introgression of archaic variants into human populations is an already known phenomenon, with some variants even providing a selective advantage such as adaptation to living in high altitudes or haplotypes carrying alleles of genes involved in the immune-system.
However, analysis of introgressed variants is restricted to SNVs (single nucleotide variants) or small indels due to degradation of ancient DNA. Here we apply a novel k-mer based approach to genotype large indels found in modern human in both Neanderthal and Denisova individuals. We test the method on present day modern humans with known genotypes (1000 genomes) and show that we find high concordance.
We genotype large indels in high coverage from data from the Altai Neanderthal, a newly sequenced Vindija neanderthal and the Altai Denisova individual and identify > 1000 variants greater than 31 bp in each archaic individual that could not be found previously. We find that shared large indels are not evenly distributed across the genome, and the highest density of variants in the HLA region (Human leukocyte antigen region). We also find regions of the genome share more structural variants with the Vindija Neanderthal than with the Altai Neanderthal or the Denisova.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA12

The independent legacy of maize diversification in South America: A test case for adaptation in dispersal using archaeogenomics

Logan Kistler [1 2,*], Fábio de Oliveira Freitas [2 3], Oliver Smith [2 4], Jazmín Ramos-Madrigal [4], Nathan Wales [4 5], Claudia Grimaldo [6], Andre Prous [7], M Thomas P Gilbert [4], Robin G Allaby [2]

[1]Department of Anthropology, Smithsonian Institution, National Museum of Natural History, Washington, DC, United States, [2]School of Life Sciences, University of Warwick, Coventry, United Kingdom, [3]Embrapa Recursos Genetics e Biotecnologia, Brasilia, Brazil, [4]Centre for Geogenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, [5]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, United States, [6]University of Oxford, Oxford, United Kingdom, [7]Department of Anthropology and Archaeology, Federal University of Minas Gerais, Belo Horizonte, Brazil

Abstract: Maize was domesticated in lowland Mexico beginning around nine millennia before the present, and became the most prominent staple crop in several regions throughout the American continents by late Prehistory. Recent ancient genome sequences from the earliest maize macro-remains have shed light on the pace and sequence of evolutionary steps toward domestication from Balsas teosinte (*Zea mays* ssp. *parviglumus*), maize's wild progenitor. However, the early dispersal of maize out of its source region with human mutualists raises questions about how the severance of teosinte admixture influenced the continued evolution of domestication traits, and how maize adapted so successfully in diverse ecological conditions. Maize appears in coastal South America by the mid-Holocene, and in the Amazon shortly after, providing a test case for the nature of adaptation during dispersal across extremely variable landscape pressures. Therefore, we sequenced nine complete genomes from archaeological maize from Andean and lowland sites from Peru, Chile, Argentina, and the eastern Brazilian savanna. Additionally, we sequenced 40 modern landraces originally collected from indigenous cultivation contexts to represent the immense modern biodiversity of South American maize, and to compare in an evolutionary context with ancient genomes. We use the resulting data to interrogate the legacy of maize's post-domestication evolution outside the context of teosinte gene flow and in diverse new environments. We evaluate gene-level selection pressures and genome-wide evolution, and we test for a role of genome restructuring in adaptation. Finally, we test specific existing hypotheses suggesting multiple independent waves of maize dispersal from Mesoamerica into South America.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA11

The Genetic History of Northern Europe

Alissa Mittnik [1 2,*], Chuan-Chao Wang [2], Saskia Pfrengle [1], Mantas Daubaras [3], Gunita Zarina [4], Fredrik Hallgren [5], Raili Allmäe [6], Vyacheslav Moiseyev [7], Valery Khartanovich [7], Anja Furtwängler [1], Aida Andrades Valtueña [2], Michal Feldman [2], Christos Economou [8], Markku Oinonen [9], Andrejs Vasks [4], Mari Tõrv [10], Oleg Balanovsky [11 12], David Reich [13 14 15], Rimantas Jankauskas [16], Wolfgang Haak [2 17], Stephan Schiffels [2], Johannes Krause [1 2]

[1]Archaeo- and Palaeogenetics Group, Institute for Archaeological Sciences, University of Tübingen, Tübingen, [2]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany, [3]Department of Archaeology, Lithuanian Institute of History, Vilnius, Lithuania, [4]Institute of Latvian History, Riga, Latvia, [5]The Cultural Heritage Foundation, Västerås, Sweden, [6]Archaeological Research Collection, Tallinn University, Tallinn, Estonia, [7]Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, St. Petersburg, Russian Federation, [8]Archaeological Research Laboratory, Stockholm University, Stockholm, Sweden, [9]Finnish Museum of Natural History - LUOMOS, University of Helsinki, Helsinki, Finland, [10]Independent researcher, Tartu, Estonia, [11]Research Centre for Medical Genetics, [12]Vavilov Institute for General Genetics, Moscow, Russian Federation, [13]Department of Genetics, [14]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, [15]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States, [16]Department of Anatomy, Histology and Anthropology, Vilnius University, Vilnius, Lithuania, [17]School of Biological Sciences, The University of Adelaide, Adelaide, Australia

Abstract: Recent genetic studies of ancient human genomes have revealed a complex population history of modern Europeans involving at least three major prehistoric migrations that were influenced by climatic conditions, the spread of technological and cultural innovations and possibly diseases. To what extent these dynamics also affected the very North of the European continent surrounding today's Baltic Sea is less well understood.

Here we report novel genome-wide DNA data from 24 ancient North Europeans ranging from ~7,500 to 200 BCE spanning the transition from a mobile hunter-gatherer to a sedentary agricultural lifestyle, as well as the adoption of bronze metallurgy. We co-analyze our data with over 300 available ancient genomes and data from around 3,800 modern individuals to show that the settlement of Scandinavia occurred via a southern and a northern route, and that the first Scandinavian Neolithic farmers derive their ancestry from Anatolia 1000 years earlier than previously demonstrated. We reveal that the range of Western European Mesolithic hunter-gatherers extended to the east of the Baltic Sea, where this population persisted without gene-flow from Central European farmers until around 2,900 BCE when the arrival of steppe pastoralists introduced a major shift in economy and established wide-reaching networks of contact within Europe during the Late Neolithic and Bronze Age. These, together with continued gene-flow from the local hunter-gatherer population led to the genetic makeup of today's Lithuanian populations, while additional admixture related to Siberian and East Asian populations is needed to explain modern Estonians' genetic composition.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-125

Metabolic changes induced by human-specific amino acid substitution in ADSL protein

Vita Stepanova [1,*], Ekaterina Khrameeva [1], Kaja Moczulska [2], Philipp Khaitovich [1], Svante Pääbo [2]

[1]Center for Data-Intensive Biomedicine and Biotechnology, Skoltech, Moscow, Russian Federation, [2]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Abstract: 96 amino acid substitutions in 87 proteins distinguish modern humans from closely related archaic relatives: Neanderthals and Denisovans. One of these substitution (A429V) occurred in adenylosuccinate lyase (ADSL) protein that catalyzes two reactions in *de-novo* purine biosynthesis pathway. ADSL deficiency was linked to mental retardation and autistic behavior.

We analyzed concentrations of 353 metabolites in nine tissues of wild type mice and transgenic mice containing humanized ADSL gene. The metabolite concentrations differed significantly between wild type and transgenic mice only in brain tissues: cortex and cerebellum. Furthermore, seven of the 45 significantly different metabolites belong to *de-novo* purine biosynthesis. Interestingly, most of purine biosynthesis metabolites decreased the concentration in humanized mice.

We further compared the average concentrations of metabolites involved in purine biosynthesis between humans and non-human primates. Notably, nine of the 11 detected metabolites had lower concentrations in human cortex and cerebellum, indicating relevance of metabolic changes detected in the mouse model to recent human evolutionary history.

Keywords: ADSL; brain metabolism; human evolution

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA13

Investigating population continuity with ancient DNA under a spatially explicit simulation framework

Jeremy Rio [1,*], Nuno Silva [1], Mathias Currat [1 2]

[1]Genetic and Evolution, Anthropology Unit, University of Geneva, [2]IGE3, Institute of Genetics and Genomics in Geneva, Geneva, Switzerland

Abstract: Advances in sequencing technologies allow retrieving ancient DNA (aDNA) from fossils, providing direct snapshots from genomic diversity in past populations. Models used to explore genetic relationships between samples from different periods usually consider panmictic populations and thus neglect the spatial dynamics of genes through time. Here, we present a new spatially explicit simulation approach using serial coalescence that simulates genomic diversity in a series of samples of different ages with a modified version of the program SPLATCHE. Our model takes into account population structure, migration and the spatiotemporal variance of lineages within ancient population samples. We apply our new approach to the study of European's prehistory. Archeological documentation indicate a progressive change in lifestyle across Europe during the Neolithic transition. In order to understand if this cultural and economic change was accompanied by a genetic replacement in central Europe, we estimate the amount of population continuity between Paleolithic hunter-gatherers and Neolithic farmers using two ancient genomes from this area. The approach commonly used so far permits the rejection or conservation of a null hypothesis of full population continuity between samples from different times. Our method improves this test by leading to a non-binary conclusion: even if we reject complete population continuity, we are able to refine this result by estimating the most probable genetic contribution of local hunter-gatherers to the final Neolithic population in central Europe. Our approach constitutes a promising tool to be applied to many species for the analysis of numerous aDNA datasets being presently produced.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-131

Estimation of Ancient Nuclear DNA Contamination Using Linkage Disequilbrium

Nathan Nakatsuka [1,*], Eadaoin Harney [1], Nick Patterson [2], David Reich [1 2]

[1]Genetics, Harvard Medical School, Boston, [2]Genetics, Broad Institute, Cambridge, United States

Abstract: Ancient DNA (aDNA) has been a revolutionary technology for inferring human history, but unfortunately, these inferences can be distorted by sample contamination. The best current method for estimating rates of contamination in aDNA using nuclear genome data focuses on detecting polymorphism on the X chromosome in males, but this does not work in females. Methods based on autosomal DNA have been developed for modern DNA, but these require genotype array data of uncontaminated samples, accurate knowledge of the sample's population allele frequencies and/or knowledge of all potential contaminant individuals' SNP genotypes, which are generally not available for aDNA.

Here we report a novel method for estimating autosomal aDNA contamination using patterns of linkage disequilibrium (LD) in the sample. Our algorithm is based on the idea that sequences from a contaminating individual will diminish the LD within the sample individual, because they are from different haplotypes. We use reference panels from 1000 Genomes populations to attain approximate background haplotype and SNP frequencies and estimate contamination by fitting a model where contamination and the expected test sample's haplotype distribution produce the observed sample's haplotype distribution. We correct for mismatch with the reference panels using damaged reads (which removes modern contaminants).

Our method accurately infers contamination generated in simulations using widely divergent 1000 Genomes populations as well as in real ancient samples, and has standard errors less than 1.5% for contamination of 5% or higher in samples with at least 400,000 snps covered.

Expanded summary*: Ancient DNA has been a revolutionary technology for inferring human history, allowing direct observation of the genetics of individuals who lived thousands of years in the past. Unfortunately, these inferences can be distorted by sample contamination during the intensive processing required to extract the DNA.

Methods for measuring contamination based on mitochondrial DNA can be biased due to the large differences between mitochondrial and nuclear DNA. The best current method for estimating rates of contamination in aDNA using nuclear genome data focuses on detecting polymorphism on the X chromosome in males, but this does not work in females. Two contamination detection methods developed in the past few years are commonly used for modern DNA. ContEst, developed by the Broad Institute, uses a Bayesian approach to calculate probability of contamination level based on the Phred-like Q-scores, number of reads covering the site, number of homozygous SNP sites, and allele frequencies (known based on a panel). This method was developed primarily for cancer genomics and requires genotype array data of uncontaminated DNA, which is generally not available for aDNA samples. ContaminationDetection, developed at University of Michigan, models the data as a mixture of two samples (independent Gaussians) to attain a maximum likelihood estimate of contamination. Unfortunately for aDNA analyses, this method requires one to have genotype information of all possible contaminant individuals beforehand, which is usually not possible for aDNA due to the large number of individuals that might process the samples over time. Lastly, both of these methods require accurate knowledge of the sample's population allele frequencies, which are often not available for aDNA. Another method developed specifically for ancient DNA incorporates the contamination into the demographic estimates. This in principle is more powerful, but it can severely limit many demographic tests that are not based on explicit population modeling or have models that are more complicated.

We developed a novel method for estimating autosomal aDNA contamination using patterns of linkage disequilibrium (LD) in the sample. Our algorithm is based on the idea that when sequences from an alternative individual (i.e. contaminant DNA) are present in a sample, the LD within the sample individual will be diminished, because the contaminant DNA is on a different haplotype block and therefore should have no LD with the authentic ancient DNA sample. We use reference panels from 1000 Genomes populations to attain approximate background haplotype and SNP frequencies and estimate contamination by fitting a maximum likelihood model

where contamination and the expected test sample's haplotype distribution produce the observed sample's haplotype distribution. We correct for mismatch with the reference panels using damaged reads (which removes modern contaminants).

Our method accurately infers contamination generated in simulations using widely divergent 1000 Genomes populations as well as in real ancient samples, and has standard errors less than 1.5% for contamination of 5% or higher in samples with at least 400,000 snps covered. This will allow users to detect samples with 5% or more contamination with high confidence so that they can be removed from analyses.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-135

Ancient Iberian and Central European Yersinia pestis genomes reveal genetic diversity of the Justinianic Plague

Marcel Keller [1 2,*], Maria A. Spyrou [1 3], Brigitte Haas-Gebhard [4], Bernd Päffgen [5], Jochen Haberstroh [6], Albert Ribera [7], Kathrin Nägele [1], Bernd Trautmann [2], Joris Peters [2 8], Alexander Herbig [1], Kirsten Bos [1], Michael McCormick [9], Michaela Harbeck [2], Johannes Krause [1]

[1]Department Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, [2]State Collection of Anthropology and Palaeoanatomy, Bavarian Natural History Collections, Munich, [3]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, [4]Archaeological Collection of the Bavarian State, [5]Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig-Maximilian University Munich, [6]Bavarian State Department of Monuments and Sites, Munich, Germany, [7]Department for Municipal Archaeology, Valencia City Council, Valencia, Spain, [8]Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig-Maximilian University Munich, Munich, Germany, [9]Initiative for the Science of the Human Past, Department of History, Harvard University, Cambridge, United States

Abstract: The Justinianic Plague (541-750 CE) was a pandemic that contributed to the end of Antiquity, probably entering the Roman Empire at the Red Sea and spreading through and beyond it. Recently molecular evidence from two Early Medieval grave fields (Aschheim & Altenerding) in SE Germany identified *Yersinia pestis* as the likely causative agent of the Justinianic Plague. Phylogenomic analysis showed the Central European 6th century Justinianic strains to form a previously unknown branch in the bacterial phylogeny, distinct from the lineage associated to the second pandemic that followed the Black Death (1348 - 1351 CE). Here, we present the first genomic data of *Y. pestis* from the western part of the Mediterranean (Valencia, Spain) as well as new *Y. pestis* genomes from multiple sites in Central Europe from the 6th to 7th century. We confirm a common origin for all Justinianic strains sequenced to date, but demonstrate novel aspects of its evolution. Additionally, the phylogenetic analysis suggests at least two independent waves of plague in Central Europe, both unknown in contemporary records. In contrast to what is known regarding the second pandemic following the Black Death, we observe the formation of multiple plague lineages that reflect different epidemiological patterns.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA10

40,000-year-old individual from Asia provides insight into early population structure in Eurasia

Melinda Yang [1 2,*], Xing Gao [1], Ayinuer Aximu Petri [1 3], Haowen Tong [2], Birgit Nickel [3], Matthias Meyer [1 3], Svante Pääbo [1 3], Janet Kelso [1 3], Qiaomei Fu [1 2]

[1]Laboratory on Molecular Paleontology, Max Planck Institute for Evolutionary Anthropology and the Institute of Vertebrate Paleontology and Paleoanthropology, [2]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Beijing, China, [3]Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Abstract: To date, very few ancient genomic studies have been conducted in Asia. Genome-wide studies using ancient individuals from Europe have revealed complex ancestry and genetic structure in ancient populations that could not be observed studying only present-day populations, suggesting similar approaches may also aid in elucidating the demographic history in Asia. Here, we present genome-wide data for a 40,000-year-old individual from Tianyuan Cave near Beijing, China. We show that he is more related to present-day Asians than present-day and ancient Europeans. However, unlike present-day Asians, he shows potential relationships with some present-day South Americans and a 35,000-year-old European individual. Our results suggest that there was extensive population structure in Asia by 40,000 years ago that persisted over an extended period of time.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-132

Evaluating the robustness of coalescent dating methods from short ancient DNA sequences

Fernando Villanea [1,*], Brian Kemp [2], Andrew Kitchen [3]

[1]IGEM, Temple University, Philadelphia, PA, [2]Anthropology, University of Oklahoma, Norman, OK, [3]Anthropology, The University of Iowa, Iowa City, IA , United States

Abstract: Molecular dating translates genetic polymorphism into a measurement of time, allowing biological events to be placed firmly in contexts of ecological and geological relevance. Molecular dating through the use of ancient DNA (aDNA) is advantageous, because calibration points are available in the form of associated radio-carbon dates. However, some aDNA data may not contain sufficient information to parameterize molecular dating models. This is due to the shorter sequence lengths typical of aDNA generated through Sanger sequencing (<200 bp). Here, we explore the robustness of molecular dating methods when applied to short aDNA sequences. We simulate two sets of sequences: the first representing contemporary DNA data, 381bp in length, sampled from the present. We then simulated a set of aDNA sequences accompanied by chronological dates, to represent the advantage of including radio-carbon dates as calibration points. The aDNA sequences are 157 bp long, limiting the analysis to a shorter alignment. We programmed a molecular rate standard in all simulations and then used BEAST to recapture this rate from the simulations. For the contemporary sequences, the estimated molecular rate was only different from the standard by 0.4% (relative to the total parameter range). For the shorter contemporary and ancient sequences, the estimated rate was only different from the standard by 2%. Our results reveal that molecular dating including short ancient sequences is surprisingly robust.

Expanded summary*: The experiment proposed in the abstract is a component of a larger study aimed at dating an ancient population expansion of northern fur seals (*Callorhinus ursinus*). The population expansion was reported originally in Dickerson and colleagues (2010) and dated to ca. 11,000 years before present. Their dating of this expansion posterior to the last glacial period (LGP) implies that northern fur seals increased their population size during a time of global warming and sea ice coverage retraction. The previous estimate, however, used a mutation rate derived from calculations of divergence time between northern fur seals and sea lions, introducing considerable uncertainty. The mutation rate used to time the population expansion in Dickerson and colleagues (2010) was 5.8E-8 substitutions/site/MY. In this study, we calculate molecular rates directly from a combination of modern and ancient DNA sequence data, and find a rate of 1.94E-07 substitutions/site/MY. Unfortunately, the ancient sequences are short (157 bp), limiting the length the alignment. We approached the disparity in estimates by considering two hypotheses; the faster mutation rate could be an artifact of short sequences, the introduction of radiocarbon dates as calibration points, and the coalescent method used for estimation (as proposed in Bandelt 2008; Debruyne and Poinar 2009; Emerson 2007). The alternative hypothesis is that rate estimation using a coalescent method is better (as supported by Ho and Shapiro 2011; Molack and colleagues 2013). In order to address the question, we use simulated sequence data to confirm the accuracy of our mutation rate estimates. The results of our experiment support the high accuracy of coalescent based estimation, even when limited to short sequences. The supported rate, in combination with improved sampling, allowed timing the population expansion to ca. 90,000 years before present (YBP), the middle LGP. The LGP, corresponding to 110,000 YBP to 12,000 YBP, was characterized by much colder global temperatures, and lower sea levels; exposing a larger coastal area. Such an early population expansion could be consistent with an increased abundance of rookery sites made possible by low sea levels, accessible for the rearing of northern fur seal pups. The use of simulation work for the validation of results is important, in particular when the results of our research will impact predictions of future responses of vulnerable species to changing ecological conditions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA2

Passenger pigeon genomes reveal the cost of natural selection for a large population

Beth Shapiro [1,*], André Elias Rodrigues Soares [2], Gemma Murray [2] and The Passenger Pigeon Genome Working Group

[1]University of California Santa Cruz, Santa Cruz, United States, [2]Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, United States

Abstract: Passenger pigeons were once the most abundant bird species in North America, but were driven to sudden extinction in the late 19th century following a period of intensive commercial harvest. Here, we compare molecular variation across 41 mitochondrial genomes and five high (>35X) coverage nuclear genomes from passenger pigeons and mitochondrial and nuclear genomes of their closest living relative, the band-tailed pigeon. Using these data, we investigate whether the evolutionary history of the passenger pigeon made them vulnerable to extinction. We find that passenger pigeons maintained a large population for tens of thousands of years prior to their extinction, rejecting the hypothesis that their extinction was facilitated by natural demographic cycles. We also find that their large population size resulted in an increased impact of both natural selection and GC-biased gene conversion, and that both were modulated by variation in the rate of recombination across their genome. In addition to shedding light on the extinction of passenger pigeons, our results comparing ancient and living genomes have broader implications for understanding what limits natural selection and diversity in real populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-129

Adaptation of the FADS gene family in Europe: Variation across time, geography and subsistence

Kaixiong Ye [1,*], Feng Gao [1], David Wang [1], Ofer Bar-Yosef [2], Alon Keinan [1]

[1]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY, [2]Department of Anthropology, Harvard University, Cambridge, MA, United States

Abstract: The *FADS* gene family encodes rate-limiting enzymes for the biosynthesis of omega-6 and omega-3 long chain polyunsaturated fatty acids (LCPUFAs). Positive selection on *FADS* genes had been reported in multiple populations (e.g. vegetarian Indians, seafood-eating Inuit), but its presence and patterns in Europeans remained elusive. Here, by bringing together ancient (aDNA) and modern DNA, we demonstrated that both before and after the advent of farming in Europe, *FADS* genes have been targeted by positive selection. However, adaptive alleles are *opposite* between these two periods. Very recent selection shaped an 85-kb-long haplotype, as revealed both by aDNA-based analysis of extreme frequency change, and by long-range haplotype- and site frequency spectrum-based selection tests in extant populations. All tests pointed to stronger selection in Southern populations. In contrast, the frequency of this recently-adaptive haplotype decreased drastically over earlier times, among pre-Neolithic hunter-gathers (~30,000 – 8,000 years ago). An aDNA-based time series selection test revealed that—during that period—positive selection had indeed increased frequencies of alleles opposite to those on the haplotype. These selection patterns are consistent with anthropological discoveries of a dietary shift with the Neolithic revolution from animal-based diets with significant aquatic contribution to plant-heavy diets and of stronger dietary reliance on plant in Southern European farmers. Thus, the adaptation of *FADS* genes in Europe varies across time and geography, probably due to varying diet and subsistence. Finally, we showed that recently-adaptive alleles are associated with higher *FADS1* expression, enhanced LCPUFAs biosynthesis and reduced risk of inflammatory bowel diseases in Europeans.

Expanded summary*: Identifying genetic adaptations to local environment, including diet, is not only essential for understanding human evolution but also critical for elucidating the effect of genetic variations on human health and disease. The fatty acid desaturase (*FADS*) gene family consists of *FADS1*, *FADS2* and *FADS3*, which evolved by gene duplication and cluster in a 100-kb genomic region. *FADS1* and *FADS2* encode rate-limiting enzymes for the biosynthesis of omega-3 and omega-6 long-chain polyunsaturated fatty acids (LCPUFAs). While LCPUFAs can be absorbed from animal-based diets, their biosynthesis from shorter-chain precursors is indispensable to compensate for their absence in plant-based diets. Positive selection on *FADS* genes has been shown in multiple human populations. Our recent study (*Molecular Biology and Evolution*; 2016) demonstrated positive selection on an insertion-deletion polymorphism in *FADS2*, which regulates *FADS1* expression, in populations traditionally residing on plant-based diets (South Asians, Africans, and some East Asians), but not in others. We further supported this hypothesis by functional association of the adaptive allele with enhanced LCPUFAs biosynthesis. This study has attracted wide attention of scientists, scientific journals, and the general media alike (leading to an Altmetric score of 731, top 0.03% of >7M papers ever tracked). In Europe, positive selection on *FADS* genes has only been reported recently in a study based on ancient DNA (aDNA), while evidence from modern DNA (mDNA) is still lacking. Moreover, possible differences in selection pressure, geographically between Northern and Southern Europe and temporally before and after the Neolithic revolution, have not been examined before.

In this study, with an imputed aDNA data set of 325 samples and multiple mDNA data sets from the 1000 Genomes Project (1000GP), UK10K, Human Genome Diversity Project (HGDP) and the Population Reference Sample (POPRES), we 1) performed an aDNA-based selection test of extreme frequency change between ancient and modern samples, confirming and refining recent positive selection signals on an 85-kb long haplotype in *FADS* genes in Europe; 2) performed multiple selection tests (iHS, nSL, and Fay and Wu's H) based solely on present-day samples, revealing consistent selection signals on the same genomic region; 3) compared signals between Northern and Southern Europe, unraveling stronger selection and higher adaptive allele/haplotype frequencies in the South; 4) observed drastic frequency decrease for the recently adaptive haplotype among pre-Neolithic European hunter-gatherers, from 32% ~30,000 years ago to being practically absent ~7,500 years ago; and 5) performed an aDNA-based time series selection test, revealing that positive selection had dramatically increased the frequency of alleles opposite to those on the haplotype.

These selection patterns are consistent with anthropological findings: 1) the Neolithic transition from hunter-gathering to farming in Europe caused a sharp dietary shift from animal-based diets with significant aquatic contribution to terrestrial plant-heavy diets; and 2) recent European farmers in the South relied more heavily on plants while their Northern counterparts consumed more aquatic and dairy products. To unravel the molecular mechanism and functional significance of positive selection on *FADS* genes in Europe, we found that recently adaptive alleles are associated with higher *FADS1* and *FADS3* expression, and lower *FADS2* expression, based on expression quantitative trait loci (eQTLs) from GTEx, and that they are significantly associated with 46 traits based on previously-published GWAS and an additional association analysis we carried out in data from the UK10K data set. Of note, recently adaptive alleles are associated with enhanced LCPUFAs biosynthesis and lower risk of inflammatory bowel diseases, consistent with their evolutionary significance for recent farmers and with the anti-inflammatory effects of many LCPUFAs-derived signaling molecules.

Our discovery of subsistence-based temporal and geographical variations of selection in Europe supports and completes the global picture of the local adaptation of *FADS* genes: positive selection on LCPUFAs-biosynthesis-enhancing alleles in populations relying heavily on plants (e.g. farmers in Europe, Africa and South Asia), but selection on biosynthesis-diminishing alleles in populations subsisting on lipid-rich marine or animal-based diets (e.g. hunter-gatherers in Greenland and pre-Neolithic Europe). These apparently complex but evolutionarily consistent patterns make the *FADS* locus an extraordinary example of human adaptation to local environment. The opposite pattern of positive selection in different dietary environments also highlights the potential and importance of personalizing one's diet to one's genome in the future nutritional practice.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-130

Fast and robust detection of ancestral selective sweeps

Xiaoheng Cheng [1,*], Cheng Xu [1], Michael DeGiorgio [2]

[1]Huck Institutes of Life Sciences, [2]Department of Biology, The Pennsylvania State University, University Park, United States

Abstract: Natural selection leaves footprints in genomes, which erode over time. The availability of large population genomic datasets has made it possible to identify past signatures of positive selection. However, few available methods are tailored for specifically detecting signals of positive selection in ancestral populations with multiple descendent populations from polymorphism data. We introduce the ancestral branch statistic (ABS), a four-population summary statistic to identify selective sweeps occurring in the ancestor of a pair of target populations. Simulations indicate that ABS performs at least as well as the complementary likelihood-based approach, 3P-CLR, under diverse selection scenarios. We also applied ABS on contemporary human genomic data to scan for genes that may have been adaptive to ancestral East Asian populations, and found both previously reported genes, including *EDAR* and *HERC1*, and a number of novel candidates. We also performed scans with ancient European genomic data to reexamine positive selection in ancestral Europeans. The *MCM6*/*LCT* cluster, and the pigmentation genes *SLC45A2* and *HERC2* are strong outliers, agreeing with previous studies. Finally, we have developed open-source software for readily computing ABS from allele count data in four populations.

Expanded summary*: Positive natural selection acts on beneficial phenotypic characters, leading to indirect effects on levels of genetic diversity in a population. As a result, alleles at affected loci will change in frequency in a population. As beneficial alleles spread in populations, nearby linked genetic diversity is also altered. This process is known as a selective sweep. These footprints of altered genetic diversity in a large genomic region around adaptive loci, are the signatures that methods typically look for to infer selection. A number of approaches have been developed to detect the signals of selective sweeps from genomic data. However, many lack power for old sweeps, and most cannot distinguish between selective events occurring in ancestral vs. descendent populations.

We developed the ancestral branch statistic (ABS) to identify sweeps specific to ancestral populations. ABS is a summary statistics that measures the amount of genetic change in a population directly ancestral to a pair of target populations. Using allele frequency data from a pair of target and a pair of reference populations, ABS calculates the length of the internal branch of the unrooted tree relating the two target and the two references on a given genomic region. The longer this branch, the more indicative it is that a sweep has occurred at this region in the population ancestral to a pair of target populations. ABS imposes few assumptions on the phylogeny, requires only allele frequency data, and can be a good alternative to the recently published model-based likelihood method, 3P-CLR (Racimo 2016). We therefore believe that our method offers a complementary approach to 3P-CLR, that will be applicable to both model and nonmodel

organism polymorphism data. We also developed open-source software, *CalcABS*, so that our method is readily accessible to future researchers.

We employed simulations to evaluate the performance of ABS and 3P-CLR under diverse selection scenarios, including partial and full sweeps on *de novo* mutations, and sweeps on standing variation. Both methods showed comparable power to detect strong or moderate selection on *de novo* mutations. Additionally, ABS displayed sensitivity only for sweeps in the immediate ancestral population, whereas 3P-CLR was unable to distinguish between such sweeps and those ancestral to all populations. ABS also slightly outperformed 3P-CLR for partial sweeps and sweeps on standing variation. Both methods responded similarly to background selection due to selective constraint and to population bottlenecks.

We applied ABS on contemporary human genomic data (1000 Genomes Project Consortium 2015) to scan for selection ancestral to East Asians. The top candidates include both previously reported genes, such as *EDAR* and *HERC1*, and a number of novel candidates. We further applied ABS on the merged dataset of ancient (Mathiesen *et al.* 2015) and contemporary Europeans to scan for selective sweeps in ancestral Europeans. We found outstanding signals in the *MCM6/LCT* cluster, responsible for lactase persistence, as well as the pigmentation genes *SLC45A2* and *HERC2*, agreeing with the hypothesis that Europeans developed light skin tones and tolerance to diets more recently than people had previously thought.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA3

Temporal genetic diversity in eastern gorillas

Tom Van Der Valk [1,*], Love Dalén [2], Katerina Guschanski [1]

[1]Animal ecology, Uppsala Univeristy, Uppsala, [2]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden

Abstract: Genetic diversity is recognized as one of the three forms of biodiversity that deserves global conservation attention, as it is linked to the species' capacity to adapt to changing environment. Low levels of genetic diversity can result from historically low population sizes and are not directly informative about species evolutionary potential. Therefore, evaluating temporal changes in genetic diversity is crucial for monitoring endangered animal populations. Here, we focus on two critically endangered eastern gorilla species, mountain and Grauer's gorillas. The distribution range of these great apes has been heavily impacted by war and instability, resulting in pronounced decline in population sizes within the last few decades. To disentangle the effects of historical demographic and recent anthropogenic events, we sequenced complete genomes of six mountain and seven Grauer's gorilla from museum-preserved specimens of up to 105 years old. Comparing our historical data to available modern genomes of the same species, we show how levels of inbreeding and genetic diversity change through time as effect of anthropogenic factors. The temporal perspective allows us to identify fine-grained genomic differences that have accumulated on a short timescale of a few generations, illuminating the genomic effects of fast population decline.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-143

Genomic analysis of Yersinia pestis from a 17th century outbreak in London

Maria A. Spyrou [1,*], Aida Andrades Valtueña [1], Elizabeth Nelson [1], Don Walker [2], Alexander Herbig [1], Johannes Krause [1], Kirsten I. Bos [1]

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany, [2]Museum of London Archaeology (MOLA), London, United Kingdom

Abstract: The Black Death (1347-1353 AD) initiated the second plague pandemic in Europe (14^{th}-18^{th} centuries AD), and its etiological agent has for long been a topic of academic debate. In recent years, next generation sequencing (NGS) technology enabled the reconstruction of *Yersinia pestis* genomes from Black Death victims in London, as well as from post-Black Death outbreaks in Ellwangen, Germany (1485-1627 cal AD) and Marseille, France (Plague of Marseille, 1720-1722 AD). Though plague is absent in Europe today, the sequencing of these historical genomes allowed the identification of a previously uncharacterized plague lineage that persisted in Europe during the second plague pandemic and seems to now be extinct. Here we present the reconstruction of five additional *Y. pestis* genomes from a 17^{th} century outbreak in London from this putatively extinct European lineage. We identify extensive genome decay along this terminal branch, most notably within regions associated with flagellum expression and chemotaxis. In addition, due to exceptional preservation of long *Y. pestis* fragments in one of the samples, we attempt a *de novo* assembly of its genome. Through these analyses we hope to better understand the ongoing processes of genome rearrangements and decay, which might have influenced the biology of *Y. pestis* during its 400 years of persistence in Europe.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA6

Reconstructing prehistoric African population structure and adaptation

Pontus Skoglund*, Alissa Mittnik, Kendra Sirak, Mateja Hajdinjak, Nadin Rohland, Swapan Mallick, Tasneem Salie, Anja Heinze, Matthias Meyer, Alexander Peltzer, Matthew Ferry, Eadaoin Harney, Megan Michel, Kristin Stewardson, Jessica Cerezo-Roman, Crissy Chiumia, Alison Crowther, Elizabeth Gomani-Chindebvu, Richard Helm, Mark Horton, Alan Morris, John Parkington, Mary E Prendergast, Raj Ramesar, Ceri Shipton, Jessica Thompson, Ruth Tibesasa, Vanessa Hayes, Svante Pääbo, Nick Patterson, Nicole Boivin, Ron Pinhasi [1], Johannes Krause, David Reich

[1]Department of Anthropology, University of Vienna, Vienna, Austria

Abstract: The population genomic landscape of Africa prior to its transformation by expansions of farmers and pastoralists is poorly understood, partly due to poor ancient DNA preservation and partly due to the deep time scale of human population history on the continent. We assembled genome-wide data from ten sub-Saharan Africans who lived in the last 4,500 years, and show that one of the most deeply divergent present-day human lineages that is today found almost exclusively in people living in southern Africa, was in the past 2,000 years also present in populations much farther north in Malawi and the Zanzibar archipelago. These results highlight the existence of an ancient genetic cline stretched over thousands of kilometers along a south-north axis. By leveraging data from ancient African genomes without ancestry from more recent into-Africa migrations, we show that western Africans today may harbor ancestry from a lineage that separated from other modern human lineages earlier than any other, including the Khoe-San of southern Africa. Finally, we use the availability of time-stratified southern African genomes to document evidence of both selective sweeps and polygenic selection that might have conferred adaptations to desert environments.

Expanded summary*: Africa is the homeland of our species, and contains within it more human genetic diversity than the rest of the world combined. However, far less is known about the prehistory of Africa than the prehistory of other parts of the world, both because of the poor preservation of ancient DNA in Africa's hot climate, and because of the disruptions of African population structure that occurred with the expansion of farming populations. Here we increase the amount of ancient DNA from Africa by a factor of 10 by taking advantage of recent advances for extracting DNA from ancient individuals. Using this first view of prehistoric African population structure, we provide evidence for a previously unknown hunter-gatherer population that once dominated East Africa, and the existence of an admixture gradient in which ancient East African foragers where in contact with southern African foragers as far north as Tanzania. In contrst, today such ancestry is restricted to the southern tip of Africa.. We also show evidence that West Africans today harbor substantial ancestry from a lineage that split from other modern humans before the lineage currently viewed as oldest (the Khoe-San of southern Africa). Finally, we reveal recent natural selection in the Khoe-San of southern Africa today that may have provided key adaptations to life in the open Kalahari desert, including genes affecting response to radiation and taste receptor loci. These results will provides the first view of prehistoric African population structure, and represent a first ancient genomic step into the deep past of humans in Africa.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-109

Evolution of Polyunsaturated Fatty Acid Metabolism: From Africa to the New World

Daniel Harris [1 2 3,*], Ingo Ruczinski [4], Heinner Guio [5], Floyd H. Chilton [6 7], Rasika A. Mathais [8], Timothy D. O'Connor [1 2 3]

[1]Institute for Genome Sciences, [2]Department of Medicine, [3]Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, [4]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, United States, [5]Laboratorio de Biología Molecular, Instituto Nacional de Salud, Lima, Peru, [6]Department of Physiology/Pharmacology, [7]Department of Cancer Biology, Wake Forest School of Medicine, Winston-Salem, [8]GeneSTAR Research Program, Johns Hopkins University, Baltimore, United States

Abstract: The conversion of dietary short to long chain polyunsaturated fatty acids (LC-PUFAs) is vital to a wide range of biological processes. Fatty acid desaturase (FADS) 1 and 2 catalyze the rate-limiting steps in the biosynthesis of LC-PUFAs. The FADS region contains two major alleles: 1) ancestral and 2) derived, with the derived allele acting more efficiently and being nearly fixed in Africa. We tested if the ancestral allele is indicative of archaic introgression or an ancient polymorphism with 3,506 genomes accounting for all continental ancestries. Native American ancestry is nearly fixed for the ancestral allele as 99.54% of Native American haplotypes have the ancestral allele. Further, the ancestral allele appears to be under selection in Native Americans using a Population Branch Statistic. Cold weather adaptation of this allele is further supported by a positive correlation between the ancestral allele frequency and the latitude of Siberian populations. The Neanderthal is more closely related to the derived while the Denisovan is an outgroup. In addition, the derived alleles have a time to the most recent common ancestor of 537,742 years ago (95% CI = 444,724-639,585), which overlaps the modern-archaic human divergence. Therefore, these results are in support of an ancient polymorphism forming prior to the divergence of these hominins. In addition, the near fixation of the ancestral allele in Native American ancestry is consistent with a coastal founding of the Americas. High access to seafood would likely increase LC-PUFAs in their diet, therefore lessening the selective advantage of the derived allele.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-139

Population genetics of the agricultural transition in Papua New Guinea

Anders Bergström [1,*], Stephen J Oppenheimer [2], Alexander J Mentzer [3], Kathryn Auckland [3], George Koki [4], William Pomat [4], Peter Siba [4], Yali Xue [1], Manjinder S Sandhu [1], Chris Tyler-Smith [1]

[1]Wellcome Trust Sanger Institute, Hinxton, [2]School of Anthropology and Museum Ethnography, [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, [4]Papua New Guinea Institute of Medical Research, Goroka, Papua New Guinea

Abstract: In the last 10ky, humans in different parts of the world have transitioned from hunter-gatherer to farming lifestyles, and genetic studies are increasingly indicating that the spread of farming, culture and languages during this period has primarily been driven by the spread of people and thus genes. Papua New Guinea (PNG) underwent its own independent agricultural transition, but it's not known if the population genetic consequences here were similar. We investigated this using genome-wide array genotypes from 381 individuals across 85 language groups, and 39 whole-genome sequences. We find that population structure in the highlands region has mostly formed only in the last 10kya and is characterized by a striking genetic divide to lowland populations and major increases in effective population size, consistent with a reshaping of genetic structure following the adoption of farming here. However, PNG differs from other parts of the world by having very strong genetic differentiation between groups, with many F_{ST} values exceeding those between major populations within continents. Ancient DNA suggests that at least in the case of west Eurasia, the current genetic homogeneity has actually been established only in the last few thousand years. The independent history of PNG then demonstrates that an agricultural transition does not necessarily lead to such a collapse of population structure. PNG, with its 850 languages and immense cultural diversity, might thus better reflect the population genetic structures that would have characterized most human societies until the very recent past.

Expanded summary*: Papua New Guinea (PNG) represents a key region in human population history, containing some of the oldest evidence of human occupation outside of Africa dating back to ~50 kya and today being the linguistically most diverse place in the world with approximately 850 languages (more than 10% of the world's total). It was also one of the handful of places in the world where humans developed agriculture and left behind the hunter-gatherer lifestyle. Genetic studies are increasingly indicating that PNG, and the whole continent of Sahul which also included Australia and Tasmania, has been isolated from the rest of world from the initial settlement until at least the last few thousand years. Its history therefore constitutes a second, independent 'replicate' of human evolution over ~50ky, allowing us to ask if the population genetic processes that unfolded here were similar to those in the rest of the world. In particular, there is an opportunity to ask if the transition from a hunter-gatherer to a farming lifestyle in PNG had the same effects on human population structure as it did elsewhere.

Most genetic studies in PNG to date have however been limited to small numbers of population samples and/or genetic markers. We have generated genotype array data (1.7 million markers) on 381 individuals from 85 different language groups from PNG, and whole genome sequences for 39 individuals. This represents the first large-scale study of the population genetic history of this part of the world.

We confirm the genetic independence of PNG, especially its interior highlands region, from the rest of the world. We find evidence for a population expansion in the highlands within the timeframe of the spread of agriculture, suggesting that similarly to other parts of the world, agriculture here spread though the movement of people, rather than just the spread of ideas. However, PNG differs in a major way from other parts of the world that have also undergone agricultural transitions, in that genetic differentiation is remarkably strong, with F_{ST} values exceeding those within e.g. all of Europe. This study thus demonstrates that while both Europe and PNG transitioned to agriculture, the former saw dramatic genetic homogenization while the latter did not. As such it is an important contribution to the emerging picture on the role of lifestyle and culture in shaping the evolutionary trajectories of human populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

OM-IA4

The genomic health of ancient hominins

Ali Berens, Taylor Cooper [1], Joseph Lachance [1,*]

[1]Georgia Institute of Technology, Atlanta, United States

Abstract: The genomes of ancient humans, Neandertals, and Denisovans contain many alleles that influence disease risks. Using genotypes at 3180 disease-associated loci, we estimated the disease burden of 147 ancient genomes. After correcting for missing data, genetic risk scores were generated for nine disease categories and the set of all combined diseases. These genetic risk scores were used to examine the effects of different types of subsistence, geography, and sample age on the genomic health of ancient individuals. On a broad scale, hereditary disease risks are similar for ancient hominins and modern-day humans, and the genomic health of ancient individuals spans the full range of what is observed in present day individuals. In addition, there is evidence that ancient pastoralists may have had healthier genomes than hunter-gatherers and agriculturalists. We also observed a temporal trend whereby genomes from the recent past are more likely to be healthier than genomes from the deep past. This calls into question the idea that modern lifestyles have caused genetic load to increase over time. Focusing on individual genomes, we find that the overall genomic health of the Altai Neandertal is worse than 97% of present day humans and that Ötzi the Tyrolean Iceman had a genetic predisposition to gastrointestinal and cardiovascular diseases. As demonstrated by this work, ancient genomes afford us new opportunities to diagnose past human health, which has previously been limited by the quality and completeness of remains.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-134

The effects of selection and demography on Neanderthal ancestry in modern humans

Martin Petr [1,*], Benjamin Vernot [1], Janet Kelso [1]

[1]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Abstract: Advances in ancient genomics have brought many insights into the evolutionary histories of anatomically modern humans (AMH) and Neanderthals, and we now know that all non-Africans today derive at least 1-2% of their ancestry from the Neanderthals. It is often suggested that purifying selection has acted against introgressed Neanderthal alleles, and such selection has been invoked to explain the depletion of introgression around genes and in conserved regions, large "deserts" depleted of Neanderthal introgression, and a decrease of Neanderthal ancestry over time observed from ancient and present-day AMH samples. It has been recently shown that such depletions in conserved regions are consistent with selection on weakly deleterious variants that had drifted to high frequencies in Neanderthals due to their small population size, but that were more efficiently selected against once they entered the larger AMH population. However, the extent and timing of this depletion are difficult to reproduce using standard models. In this study, we used population genetic simulations to fit a set of selection parameters and demographic models that can produce the dynamics of Neanderthal ancestry changes observed in early modern humans and present-day Europeans.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-137

Steps to the phylogenetic resolution of species-rich genera: A clear SINE

Michaela Halsey [1,*], Laura Blanco-Berdugo [1], Nicole Paulat [1], Roy Platt II [1], David Ray [1]

[1]Biological Sciences, Texas Tech University, Lubbock, United States

Abstract: Rodents are the most successful mammals on the planet in terms of species richness. Members of the genus *Dipodomys* (the kangaroo rats) occupy a wide range of habitats, from the mountains to the coast. Despite this habitat diversity, we lack a complete understanding of the relationship among members of the genus. Therefore, it is no small wonder that the complete phylogenetic status of the kangaroo rats remains unresolved. It is known that transposable elements (TEs) can influence genomic structure and may play a role in speciation. Certain TEs, such as short interspersed elements (SINEs), are useful for phylogenetic analysis. Using software that classifies repeat subfamilies based on small co-segregating regions, we determined candidate SINE subfamilies for use in a phylogenetic investigation on all 22 members of the genus *Dipodomys*, some of which are of conservation concern. We intend to employ a mobile element scanning (ME-Scan) approach, where we will construct new phylogenies to either support or refute existing hypotheses, many of which are based on older molecular techniques. The precision in data acquired from genomic sequences can finally answer - or at least illuminate - untold questions in the fields of conservation, molecular and evolutionary biology.

Expanded summary*: The kangaroo rats are, needless to say, an interesting group whose complete phylogeny has never been constructed. Many current phylogenies have been constructed based only on morphological data or with older molecular data. With next-generation sequencing now on the scene, it is possible to revisit many of these phylogenies. Here, we use short interpsersed elements (or SINEs) as molecular markers for the phylogenetic analysis of the genus *Dipodomys*. We intend to collect and use DNA from musuem specimens, which has become an important tool for molecular studies. Few members of the genus are of conservation concern and a comprehensive phylogeny can augment our understanding of this species-rich genus.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Intergrating ancient and modern DNA

POA-136

Detecting domestication phenotypes under selection in ancient maize

Markus Stetter [1,*], Jeffrey Ross-Ibarra [1]

[1]Plant Sciences, UC Davis, Davis, United States

Abstract: The domestication of maize began almost 9,000 years ago in Central Mexico and led to significant changes in the appearance of the plant. In contrast to maize, its wild ancestor teosinte shatters its seeds and has small hard kernels. Another striking difference is the branching of teosinte that has been lost during the domestication of maize. These and other phenotypic changes led to the expansion of maize around the world and made maize one of the most important crops today. Such morphological change likely took hundreds or even thousands of years. While phenotypes of ancient samples are mostly lost, DNA sequences from archaeological excavations allow observation of genotypes of maize during the time course of domestication. Testing selection on phenotypes in ancient samples is challenging because the data recovered from ancient DNA is often incomplete and only few samples per population exist. We use effect sizes from large modern maize GWAS panels to study phenotypes under selection in ancient samples from the center of domestication and the extension zone of maize. The diverse reference panel of modern maize has been sequenced and evaluated for over 50 traits, including several domestication traits and a variety of improvement traits. The increasing availability of ancient DNA sequences increases the statistical power of our approach. Studying ancient samples from different time periods during domestication might allow to distinguish domestication traits and traits under selection during post-domestication adaptation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-229

The deep origin and recent loss of venom toxin genes in rattlesnakes

Noah Dowell [1,*], Matt Giorgianni [1], Sean Carroll [1]

[1]University of Wisconsin - HHMI, Madison, United States

Abstract: The genetic origin of novel traits is a central, but challenging puzzle in evolutionary biology. Among snakes, phospholipase A2 (PLA_2)-related toxins have evolved in different lineages to function as potent neurotoxins, myotoxins, or hemotoxins. Here, we traced the genomic origin and evolution of PLA_2 toxins by examining PLA_2 gene number, organization, and expression in both neurotoxic and non-neurotoxic rattlesnakes. We found that even though most North American rattlesnakes do not produce neurotoxins, the genes of a specialized heterodimeric neurotoxin predate the origin of rattlesnakes and were present in their last common ancestor [~ 22 mya]. The neurotoxin genes were then deleted independently in the lineages leading to the Western Diamondback (*Crotalus atrox)* and Eastern Diamondback (*C. adamanteus*) rattlesnakes [~6 mya], while a PLA_2 myotoxin gene retained in *C. atrox* was deleted from the neurotoxic Mojave rattlesnake (*C. scutulatus*; ~4 mya). The rapid evolution of PLA_2 gene number appears to be due to transposon invasion that provided a template for non-allelic homologous recombination.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-228

An integrated model of phenotypic trait changes and site-specific sequence evolution

Eli Levy Karin [1,*], Susann Wicke [2], Tal Pupko [1], Itay Mayrose [1]

[1]Tel Aviv University, Tel Aviv, Israel, [2]Institute for Evolution and Biodiversity, Muenster, Germany

Abstract: Recent years have seen a constant rise in the availability of trait data, including morphological features, ecological preferences, and life history characteristics. These phenotypic data provide means to associate genomic regions with phenotypic attributes, thus allowing the identification of phenotypic traits associated with the rate of genome and sequence evolution. However, inference methodologies that analyze sequence and phenotypic data in a unified statistical framework are still scarce. Here, we present TraitRateProp, a probabilistic method that allows testing whether the rate of sequence evolution is associated with a binary phenotypic character trait. The method further allows the detection of specific sequence sites whose evolutionary rate is most noticeably affected following the character transition, suggesting a shift in functional/structural constraints. TraitRateProp is first evaluated in simulations and then applied to study the evolutionary process of plastid plant genomes upon a transition to a heterotrophic lifestyle. To this end, we analyze 25 plastid genes across 85 orchid species, spanning different lifestyles and representing different genera in this large family of flowering plants. Our results indicate higher evolutionary rates following repeated transitions to a heterotrophic lifestyle in all but four of the loci analyzed.

Expanded summary*: Dear committee,

Understanding species adaptation at the molecular level has long been a central goal of evolutionary biology and genomics research. Whole organism phenotypes, such as the ability to photosynthesize, a reproductive strategy or pathogenicity, often rise due to the function of numerous genes. Thus, shifts in species characteristics coincide with changes in the selective regime acting on the underlying genes. Therefore, examining the evolutionary process of genes in light of changes in the phenotypic state allows the detection of genes associated with the examined phenotypic trait.

We recently developed TraitRateProp, a method to study associations between a discrete phenotypic trait and the rate of sequence evolution. TraitRateProp combines models of sequence and character-state evolution such that rates of sequence evolution depend on the character state of a lineage at each point in time. Using a mixture model, TraitRateProp allows each sequence position to evolve either in a phenotype-dependent or phenotype-independent manner. Furthermore, the detection of specific sites associated with the phenotype is made possible by considering the relative fit of the trait-dependent and trait-independent categories. Using a simulation study we evaluated the performance of TraitRateProp and then applied it to study the evolutionary patterns exhibited by orchid plants that transition from a photo-autotrophic to a heterotrophic lifestyle.

Currently, we are in the last stages of developing a user-friendly web server that implements TraitRateProp and augments it with additional layers of biological information, such as 3D protein structure. I believe this web server will be available for the community in the upcoming months.

As TraitRateProp focuses on studying relationships between whole organism phenotypes and sequence evolution and as it is general and applicative to various fields of research, I believe it is of high relevance to the “Mechanisms of phenotypic evolution” symposium and its attendance. I would be profoundly grateful if the committee were to find me eligible for a student travel grant to present my research.

Sincerely,
Eli Levy Karin.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP4

Elucidating the mechanisms driving immune response in Drosophila through the analyses of transposable element-induced mutations

Anna Ullastres*, Josefa González

Abstract: While there are many SNP-based studies trying to elucidate the genetic basis of genotype-phenotype relationships, the role of transposable element (TE)-induced mutations is understudied. Recent evidences demonstrate that TEs are a powerful tool to identify the genetic basis of adaptive phenotypic traits. *Drosophila melanogaster* is a good model to study adaptation because it is original from tropical Africa and only recently colonized out-of-Africa environments. One of the most relevant traits in the colonization of new environments is immune response. In this work, we use NGS data from different *D. melanogaster* natural populations to identify TEs likely to be involved in immune response. Starting from 1,636 TEs, we identified 17 TEs that were present at high frequencies in natural populations and located nearby immune response genes. We combined allele-specific expression (ASE), enhancer assays, and RACE experiments to characterize the impact of these TEs in oral immune response to the gram-negative bacteria *Pseudomonas entomophila*. We show that the allele with the TE was differently expressed in 15 out of the 16 analyzed genes under control and/or infected conditions in at least one of the two genetic backgrounds analyzed. We also show that different TEs alter gene expression through different molecular mechanisms. Overall, our results illustrate that TEs are a good tool to bridge the gap between genotypic and phenotypic evolution.

Expanded summary*: My PhD focuses on understanding adaptation and more specifically the role of transposable elements (TEs) during this process in the model species *Drosophila melanogaster*. The process of adaptation still holds important unanswered questions: What and how many mutations are needed to produce an adaptive phenotype? What are the traits under selection during this process? What is the role of epistasis and pleiotropy in adaptation? In our lab, we try to answer these key questions by studying TE-induced mutations in *D.melanogaster* natural populations. TEs are present in almost all sequenced organisms and they can generate multiple types of mutations. We are currently characterizing several TE mutations by performing evolutionary, molecular, and phenotypic assays to elucidate their impact on phenotype and fitness. We aim to widen our understanding on the role of TEs on shaping phenotypic adaptation to new environments by following two different strategies: locus-specific and trait-specific. First, we have characterized a previously identified adaptive TE mutation and mapped it to its ecologically relevant phenotype: faster developmental time. I am the first author in this work that was published in 2015 in the journal Molecular Biology and Evolution. Second, we are studying the impact of several TE mutations in a highly conserved and ecologically relevant pathway: immune response.

The results I would like to present at SMBE are related to the second strategy. We have analyzed four different populations from Africa, North America, and Europe in order to detect candidate adaptive TE mutations. We identified a total of 58 candidate TEs for out-of-Africa adaptation, and 32 candidate TEs for adaptation in both Africa and out-of-Africa environments. Interestingly, the genes associated to these TEs are mainly involved in xenobiotic stress response, immune response, and behavior. We decided to focus on understanding the role of TEs in immune response. To do that, we have analyzed a total of 14 TEs located nearby 16 genes associated with this trait. We analyzed the expression of these genes in flies infected with the gram-negative bacteria *Pseudomonas entomophila* by using allele specific expression in F_1 hybrids. We further analyzed the candidate TEs combining enhancer assays and RACE experiments to elucidate the molecular mechanisms. Overall, we found that different TEs alter gene expression through different molecular mechanisms.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-235

Molecular evolution of morphological novelty under sexual conflict: the exemple of the key developmental gene distal-less in water strider

Augustin Le Bouquin [1,*], Abderrahman Khila [1], Locke Rowe [2]

[1]Institut de Géomique Fonctionnelle de Lyon, Université de Lyon, Ecole Normale Spérieure de Lyon, Lyon, France,

[2]Ecology and Evolutionnary Biology, University of Toronto, Toronto, Canada

Abstract: Recent work has shown that sexual conflict can be a powerful force that can drive lineage diversification and speciation. For example, male water striders (Hemiptera: Gerridae) of genus *Rheumatobates*, display large amounts of diversity in grasping appendages used for grasping females during premating struggles. However, although sexual conflict is thought to be a driver of diversification of these morphologies, the developmental mechanisms underlying these evolutionary changes remain poorly understood. Our lab has previously shown that the gene *distal-less (dll)* is necessary for the complete development of the elaboration of male antennae in the species *Rheumatobates rileyi*. In particular, *dll* knock-down males are viable but fail to grasp resistant females. Here, we have shown that *dll* has the same role in two sister species (*R. tenuipes* and *R. truliger*). In addition, we have found that *dll* in *R. rileyi* acquired new isoforms that water striders outside the genus *Rheumatobates* don't harbor. Preliminary results suggest that the new isoforms are sufficient for the proper development of the antennal elaboration. Further work will focus on the role of *dll*, in male antennal modification, in more distantly related species where these modifications have risen independently. Other work will test whether these isoforms are sex specific and whether they are present in all species where *dll* is shown to have a role in male antennae elaboration. This work helps to further our understanding of how changes in developmental processes can drive phenotypic divergence between the sexes within as well as between species in the context of sexual conflict.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-271

Genetic link between body pigmentation and cuticle features in Drosophila

Noriyoshi Akiyama [1,*], Aya Takahashi [1 2]

[1]Department of biological sciences, [2]Research Center for Genomics and Bioinformatics, Tokyo Metropolitan University, Hachioji-shi, Tokyo, Japan

Abstract: Natural variation of body pigmentation intensity in *D. melanogaster* is present in many different populations. Previous studies have shown positive correlations between pigmentation intensity and altitude or latitude of origin in populations from different continents. These clinal patterns suggest that pigmentation traits may be under some form of nature selection in this species. One possible selective pressure may be related to desiccation tolerance, which has been one of the environmental factors suspected to be associated with body pigmentation. However, there have been conflicting views on whether the association actually exists in this species. Thus, in order to experimentally test the effect of pigmentation associated genes, we focused on the genes involved in melanin biosynthesis pathway and conducted genetic manipulation to investigate direct effects of their expression levels on desiccation tolerance traits. Our results indicated that knocking-down or over-expressing, *ebony*, which is the major locus controlling pigmentation variation, affects dehydration speed in low humidity, but not unidirectionally. The relationship may not be straightforward, and it may partly explain the inconsistency of the association among previous studies. We also compared the body pigmentation intensity with the published cuticular hydrocarbon component profiles using the *Drosophila* Genetic Reference Panel (DGRP). Pigmentation traits showed an association with some cuticular hydrocarbon, which suggests a genetic link between pigmentation intensity and some cuticle features related to water permeability. Taken together, widely observed variation in body pigmentation intensity in this species may be reflecting a complex genetic association between cuticle features and physiological traits such as desiccation tolerance.

Expanded summary*: Insect body color is considered to be reflecting habitat adaptation, mimicry and sexual selection in many cases, and has been paid attention as a model case of adaptive evolution. Elucidation of the molecular mechanisms underlying adaptation is a major proposition of evolutionary biology. However, studies of the molecular basis of adaptive traits are limited because of the difficulty in conducting genetic manipulation. The greatest advantage of using *D. melanogaster* in this study is the availability of the tools to manipulate the expression level of genes using transgenic flies. Several studies reported relevance between body color traits of this species and environmental stress tolerance (desiccation, low temperature, UV, etc), but there is no example analyzing molecular mechanism at the gene level. Therefore, we focused on the genes involved in body color formation and conducted a genetic manipulation to investigate direct effects of their expression levels on desiccation tolerance traits. Moreover, this study is expected to link molecular evolutionary research and ecological study when mechanism underlying desiccation tolerance and body color is elucidated. In addition to genetic manipulation experiment, we also compare body color traits with cuticular hydrocarbon components in a wild-derived population, which is related to water permeability of the cuticle.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-233

Evidence of genetic adaptation in frequently vectored yeasts

Sarah L. Seagrave [1,*], Geetanjali Minsky [2], Lilla Bartkó [2], Taom Sakal [2], Kelly M. Thomasson [2], Stephen R. Proulx [2]

[1]Molecular, Cellular and Developmental Biology, [2]Ecology, Evolution and Marine Biology, University of California, Santa Barbara, United States

Abstract: Microbes can be dispersed by a range of insects and gain access to new habitats. Dispersal through gut-vectoring involves strong selective pressures due to the caustic environment of the insect's digestive tract. Recent studies have suggested that sporulation contributes to survival of the yeast –*Saccharomyces* cerevisiae– during ingestion and digestion by insects. We analyzed key functional genes in multiple yeast populations after multi-generational exposure to the digestive tract of the fruit fly, *Drosophila melanogaster*, in a long-term evolution experiment. We determined which new sequences were observed in multiple replicate populations, as distinct from the those of the ancestral yeast genome. These were also compared to a control population which had experienced serial transfer without digestion. These results may help to increase our understanding of how genetics and epigenetics contribute to phenotypic change. Based on these results, we will engineer the derived mutations into the ancestral strain to determine the functional basis of these mutations.

Statement: I joined an Evolutionary Biology lab to answer questions I've always wondered about, but instead, I have more questions now than I had in the beginning. This is why I know I want to go into research. Beyond my four years of independent research and laboratory volunteering, I have spent five years studying Microbiology under the department of Molecular Cellular and Developmental Biology at the University of California Santa Barbara. Once I am granted my undergraduate degree, I intend to go on to get my PhD in Synthetic Biology. I hope to strengthen my inspiration for this aspiration at the SMBE conference. I'm ecstatic about talking to many bright and achieved minds in the field of Molecular Biology and Evolution so that I can hear their discoveries and ideas and see what they think of mine. I'm also hoping to get a glimpse at what my future holds if I work hard enough. Mostly, I am hoping to experience the excitement of so many like-minded people all in one place. Although I've presented two posters at symposiums funded by the University of California, I've never been to a conference of this size. When I ask what it is like, a graduate student in my lab jokes that this conference will be like my Disneyland, and I can't wait.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP5

Phenotypic adaptation by positive selection on long-term balanced alleles

Philip Kleinert [1,*], Cesare de Filippo [2], Silvia Ghirotto [3], Aida Andres [2]

[1]Bioinformatics, Max-Planck-Institut for Molecular Genetics, Berlin, [2]Evolutionary Genetics, Max-Planck-Institut for Evolutionary Anthropology, Leipzig, Germany, [3]Department of Life Science and Biotechnology, University of Ferrara, Ferrara, Italy

Abstract: Phenotypic adaption is critical for the survival of populations in changing environments, and is largely mediated by positive selection on phenotypically relevant alleles. Under classical models of positive selection the availability of adaptive alleles limits adaptation in populations of low effective population size. However, even in small populations loci under long-term balancing selection contain high levels of polymorphism and alleles that affect both phenotype and fitness. Balanced alleles hold thus the potential to mediate fast, adaptive phenotypic change upon environmental change. Identifying shifts in selective pressure across populations is challenging, so we developed a new, powerful statistic (DIFFSS) that, combined with Approximate Bayesian Computation, has high power to identify positive selection on previously balanced loci. We applied our statistics to the 1000 Genomes dataset to detect instances of novel adaptation during the first colonization of non-African environments. We identified 308 genes with significant signatures of long-term balancing selection in Africa and subsequent positive selection in Eurasia. The list includes the three previously reported cases (*PKDREJ, VKR3, SDR39U1*) and is both enriched in Keratin Filaments and depleted of MHC-Region genes. One interesting exception is *HLA-DPA1*, which shows significant signatures of local adaptation in the Finnish population. We also observe strong signatures of local adaptation in Europe in *IFI16*, a gene encoding a protein involved in viral DNA degradation in the nucleus. Our results show that positive selection on previously balanced alleles is a mechanism that mediates phenotypic adaptation to novel environments in human populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-226

Character evolution and species diversification in the Impatiens genus

Sudhindra Gadagkar*, Alisha Harrison [1], Scott Soby [1]

[1]Biomedical Sciences, Midwestern University, Glendale, United States

Abstract: The plant family Balsaminaceae comprises two genera: Impatiens and Hydrocera. Interestingly, while Impatiens is highly speciose (with at least 1000 known species, and several more added every year) the only sister genus, Hydrocera, has one solitary species, H. triflora. Additionally, there is enormous diversity in phenotypic traits among the various Impatiens species, making this genus an excellent model for understanding the association between trait evolution and speciation. We collected tissue samples from 56 Impatiens species and sequenced five markers, namely rbcL, matK, psbA-trnH, ITS2, and the atp spacer, for phylogenetic inference, and trait data on these species for 80 traits. We present the results linking character evolution to species diversification based on analyzing these trait data on the phylogeny of the 56 Impatiens species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP7

Probing the determinants of fitness effects using transposon insertion sequencing in bacteria

Alejandro Couce [1,*], Alexandra Baron [1], Olivier Tenaillon [1]

[1]French Institute of Health and Medical Research (INSERM), Paris, France

Abstract: What determines the effect of a mutation? Few topics have attracted as much attention in evolutionary genetics - and yet remain so poorly understood. At least two systematic patterns have emerged from experimental efforts over the last decade. The first pattern is the rule of declining adaptability, whereby adaptive mutations tend to be less beneficial in fitter genotypes. A second trend is that mutations become on average more beneficial under stressful conditions. Most studies so far have been limited by their small sample size or by a lack of characterization of the mutants employed. Here we circumvented these limitations by analyzing high-throughput sequencing data of transposon mutant libraries generated on bacteria from Lenski's Long-Term Evolution Experiment. In a first set of experiments, we found that the genome-wide distribution of fitness effects becomes depleted in beneficial mutations in later generations. This observation supports a modular epistasis model of adaptive evolution, in which the rule of declining adaptability emerges without the need for any epistasis sensu strictu at the genetic level. However, we also identified some instances of diminishing returns epistasis affecting particular genes, suggesting that both types of epistasis should be taken into account to produce a successful theory. Lastly, we estimated the distribution of fitness effects across environments differing in resource availability. While we recover a correlation between fitness and environmental quality, we uncover also a significant degree of idiosyncrasy that can be traced back to the specific aspects of E. coli physiology in each medium.

Expanded summary*: Mutations are the raw material for adaptive evolution, so much effort has been devoted to understanding what determines their effects on fitness. Characterizing the distribution of fitness effects (DFE) of spontaneous mutations is central to many fundamental evolutionary theories, including the origin of sex, the evolution of ploidy or the extinction of small populations. A large body of work over the past decade has unveiled that the DFE exhibits, in general, an exponential-like shape, with the vast majority of mutations being neutral and deleterious and just a tiny fraction being substantially beneficial. More recently, researchers have focused on identifying the major factors that can change the properties of the DFE. Prominent among these are the level of adaptation to a particular environment, and the degree of stringency that the environment imposes upon the organisms's physiology. Regarding the first factor, evidence has accumulated in favour of the so-called rule of declining adaptability, that is, that less fit genotypes adapt markedly faster than their fitter counterparts. Such a fact implies some kind of epistasis among beneficial mutations (i.e. their combined effects on fitness are non-trivial), which has profound implications for the repeatability of the evolutionary process. On the other hand, several studies have shown that environmental stress increases the fraction of mutations that are beneficial, as well as reduces the effect-size of deleterious ones. However, the insights gained from these studies have been limited by their relatively low throughput and the absence or incompleteness of genetic characterization of the mutants analyzed.

Here we circumvented these limitations using Insertion Sequencing (INSeq) transposon mutant libraries generated on bacteria from Lenski's Long-Term Evolution Experiment (LTEE). The INSeq technology affords the means to track the frequency of an extensive genome-wide variety of insertion mutants during competition experiments, hence enabling us to estimate the fitness effect of disrupting each and every gene of the bacteria. In a first set of experiments, we found that the genome-wide distribution of fitness effects becomes depleted in beneficial mutations in bacteria isolated from later generations. This observation supports a modular epistasis model of adaptive evolution, in which the rule of declining adaptability emerges without the need for any epistasis sensu strictu at the genetic level (microscopic epistasis). Under this view, most of the beneficial mutations are redundant and can be grouped according to the functional module they affect (e.g., altering the flux of central carbon metabolism or the usage of cytoplasmic reserves). Just a handful of mutations might be enough to improve the performance of one of these higher-level modules, rendering many other originally beneficial mutations effectively neutral. However, we also identified some instances of diminishing returns epistasis affecting particular genes, suggesting that both modular and microscopic epistasis should be taken into account to produce a successful theory. Lastly, we estimated the distribution of fitness effects across environments differing in resource availability. While we recover a correlation between fitness and environmental quality, we also uncover a significant degree of idiosyncrasy that can be traced back to the specific aspects of E. coli physiology in each medium. Overall, our results provide the highest resolution to date of

the determinants of mutation fitness effects, potentially helping to address key issues in evolutionary theory, from fixation dynamics to the genetic architecture of organisms. Moreover, the general patterns recovered here support the intriguing prospect that evolution might perhaps be predictable at the macroscopic level.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-227

The ancestral transcriptome and the evolution of sexual phenotypes in filamentous fungi

Jeffrey Townsend [1,*], Zheng Wang [1], Kayla Stefanko [2], Caitlyn Cubba [2], Frances Trail [2]

[1]Biostatistics, Yale University, New Haven, [2]Plant Biology, Michigan State University, East Lansing, United States

Abstract: Changes in gene expression have been hypothesized to play an important role in the evolution of divergent morphologies. To test this hypothesis in a model system, we examined differences in fruiting body morphology of five filamentous fungi, culturing them in a common garden environment and profiling genome-wide gene expression at five developmental stages. We reconstructed ancestral gene expression, identifying genes with the largest evolved increases in gene expression across development. Conducting knockouts and performing phenotypic analysis in two divergent species typically demonstrated altered fruiting body development in the species that had evolved increased expression. Our evolutionary approach to finding relevant genes proved far more efficient than other gene deletion studies targeting whole genomes or gene families. Combining gene expression measurements with knockout phenotypes facilitated the refinement of Bayesian networks of the genes underlying fruiting body development, regulation of which is one of the least understood processes of multicellular development.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP11

Selection on polygenic traits

Laura Hayward [1,*], Guy Sella [2]

[1]Mathematics, [2]Biological Sciences, Columbia university, New york, United States

Abstract: A major objective in evolutionary biology is to uncover the genetic basis of adaption. Evidence suggests that the classic model of adaption --- as proceeding via successive advantageous alleles arising in the population and sweeping swiftly to fixation --- has not been the main mode of recent adaption in humans. Many traits of interest, in humans and other species, are highly polygenic and there is good reason to believe recent adaption in humans is mostly polygenic. To better understand the dynamics of polygenic adaptation, we consider the selective response of a polygenic trait initially at equilibrium subject to a sudden shift in fitness optimum. We use Monte Carlo simulations and mathematical modeling to better understand how the adaptive response to a change in selective pressures depends on the architecture of the trait (i.e. the number of segregating variants affecting the phenotype, and their effect sizes and frequencies). We find relationships between the evolutionary parameters underlying a trait (e.g. mutation rate, distribution of effect sizes, etc) and its adaptive response, both at the level of individual variants and at the "macroscopic" level (e.g. the trajectories of the mean and variance of the system after the shift) --- where there turns out to be an unexpected role for the third moment of the trait distribution, especially when the shift is large and the initial trait variance is small.

Expanded summary*: An important question in evolutionary biology is the extent to which the obvious phenotypic differences between human populations are due to selection, as opposed to drift or other demographic forces. In particular, we would like to understand how evolutionary forces combine to shape adaptive responses. The classical explanation, offered by the recurrent sweep model --- whereby adaption takes place in intermittent steps via successive advantageous alleles arising in the population and sweeping swiftly to fixation --- has not been borne out by the analysis whole genome data (Coop, 2009)(Hernandez, 2011). The sweep model is premised on the assumption that the trait in question is controlled by one or a few loci of relatively large effect size, but in reality multiple lines of evidence suggest that many traits of interest, both in humans and other species, are highly polygenic (i.e. controlled by many loci on the genome, each of small effect size).

In general statistical tests aiming to detect signatures of selection in genome-wide sequences have been designed with sweep models in mind and, hence, are mostly good at detecting strong sustained selection at candidate loci, but lack power to detect small recent shifts in large numbers of alleles (Pritchard, 2010). In order to design tests to detect selection on quantitative traits it would be useful to have a theoretical framework linking the underlying genetic architecture of a trait (i.e. the number of segregating variants affecting the phenotype, and their effect sizes and frequencies) to its selective responses. Not only would such a model provide us with a better understanding of the evolutionary forces that shape selective responses of quantitative traits, which is interesting in its own right, but it may also help to answer natural questions arising from observations of selection on quantitative traits, both in breeding experiments and in the wild. Of particular relevance is the hope that, in analogy to sweep models leading to tests for monogenic adaption, a better theoretical understanding of the response of highly polygenic traits to selection could be used to design tests more suited to detect selection on quantitative traits than those currently available.

Recent progress has been made in understanding how the genetic architecture of a quantitative trait at equilibrium - that is, a polygenic trait in an unchanging environment with population mean at the fitness optimum - arise from population genetic parameters (Simons and Sella, in prep.). We extend this model to provide a theoretical framework for the case of a polygenic trait, initially at equilibrium, subject to a sudden shift in fitness optimum. We use both simulations and theory to find relationships between the evolutionary parameters underlying a trait (e.g. mutation rate, distribution of effect sizes, etc) and its adaptive response.

We find there are three dynamic phases to the adaptive response. During the first, the distance of the population mean from the new optimum swiftly declines and the phenotypic variance soars. Individual variants are subject mostly to directional selection. During the very long final phase the average distance from the optimum gradually decays, from slightly positive to zero, and the variance slowly returns to its equilibrium value. Individual variants are subject to variance reducing selection and the distribution of variants frequencies and effect sizes gradually returns to the equilibrium state. During the short intermediate phase the effects of directional selection and variance reducing selection are comparable.

Surprisingly we have found in Phase I the third moment of the trait distribution spikes significantly and plays an important role in the dynamics of the adaptive response. For example, it is this skewness in the trait distribution that prevents the average mean phenotype from quite reaching the new optimum until many generations after the shift. We map the effects of different evolutionary parameters (e.g. population size, mutation rate) --- particularly that of the distribution of variants' effect sizes --- on different aspects of the adaptive response, such as the distribution of variants from standing variation that ultimately contribute to adaption.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-217

Identifying alternative genetic paths and biochemical mechanism by which historical changes in molecular functions could have evolved

Brian Metzger [1,*], Joe Thornton [2]

[1]Ecology and Evolution, [2]Ecology and Evolution; Human Genetics, University of Chicago, Chicago, United States

Abstract: Recent work dissecting the evolution of novel molecular functions has suggested that the genetic paths and biochemical mechanisms by which new functions evolve can often be quite simple. But why did historical functions evolve by the paths and mechanisms that they did? Was the historical path simply one of many possible routes, or were alternative paths and mechanisms not available? Was the outcome of evolution contingent on low probability events, and thus unpredictable, or can we predict the paths and mechanisms evolution is most likely to take? Answering these questions requires determining not only what happened historically during evolution, but also what else was evolutionarily available. Here we describe recent work combining ancestral protein reconstruction, high throughput experimental evolution, and deep mutational scanning to address these questions. We focus on the steroid hormone receptors, a family of DNA binding proteins that have undergone a functional shift in DNA specificity following a historical duplication. By evolving the DNA binding domain of ancestral steroid receptors towards their historically derived DNA specificity, we can estimate the repeatability of the evolutionary process, the set of alternative paths and mechanism that were historically available to evolution, and thus the probability that evolution would have taken the historical path. In addition, by applying deep mutational scanning to steroid receptor DNA binding domains with ancestral and derived DNA specificity, we can determine how site-specific amino acids preferences have changed over evolutionary time and following historical transitions in function. Together, these approaches allow us to 'replay the tape of life' and place historical changes in function within the broader context in which they evolved.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP10

Rapid Evolution of Lighter Skin Pigmentation in Southern Africa

Meng Lin [1 2,*], Rebecca Siford [2], Alicia Martin [3 4 5], Marlo Möller [6], Eileen Hoal [6], Marcus Feldman [7], Carlos Bustamante [3], Christopher Gignoux [3], Brenna Henn [2]

[1]Graduate Program in Genetics, [2]Dept. of Ecology and Evolution, SUNY Stony Brook, Stony Brook, NY, [3]Dept. of Genetics, Stanford University, Stanford, CA, [4]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, [5]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, United States, [6]DST/NRF Centre of Excellence in Biomedical Tuberculosis Research,S A MRC Centre for Tuberculosis Research Division of Molecular Biology and Human Genetics, Stellenbosch University, Tygerberg, South Africa, [7]Dept. of Biological Sciences, Stanford University, Stanford, CA, United States

Abstract: Skin pigmentation is under strong directional selection, with lighter skin in northern European and East Asian populations, and darker skin in equatorial populations. However, selection on skin color and its mechanisms have only rarely been elucidated in studies of other populations worldwide. KhoeSan populations in far southern Africa, who are among the earliest diverged human populations, possess lightened skin pigmentation. We sequenced pigmentation genes to high coverage in over 400 KhoeSan individuals and demonstrate that a canonical skin pigmentation gene, *SLC24A5*, experienced recent adaptive evolution in the KhoeSan. The functionally causative skin lightening allele is present at a high frequency of 24% in the KhoeSan, after controlling for the recent European gene flow. The effect size of the allele is slightly larger than the mean pigmentation difference between Europeans and East Asians, explaining 11.9% of the variance in pigmentation in the KhoeSan. Haplotype analysis indicates that the derived haplotypes in these populations are identical to those fixed in Europeans. Using a hidden Markov model, we estimate the age of the ancestral haplotype carrying the derived allele in KhoeSan to be 18 kya [12 kya – 42 kya], somewhat older than the age of the allele in Europeans at 13 kya [6 kya – 41 kya]. We hypothesize that the allele was only introduced into the KhoeSan within the past 3,000 years, likely by pastoralists moving from eastern Africa to southern Africa while retaining non-African admixture. We test this hypothesis using an approximate Bayesian computation (ABC) approach, incorporating demographic models and selection. The *SLC24A5* locus is a rare example of strong, ongoing, parallel adaptation adopted through gene flow in recent human history. We demonstrate a novel strategy of tracing the selection on both the genotype and corresponding phenotype, by modeling the signal from the genetic association and its selection through demographic history.

Expanded summary*: Human skin pigmentation is among the most notably diverse phenotypes across populations. It's also one of the most strongly selected phenotypes in the recent human adaptation history, mirroring the migration paths to different latitudes where ultraviolet radiation (UVR) varies. To present, the genetic basis and evolution of skin pigmentation have been primarily studied in light skinned northern Europeans and East Asians, mostly discovered through strong signals in selection scans. The evolution mechanism of skin color in other parts of the world remains a huge mystery.

In this study, we explore a rapid adaptation scenario of a skin lightening allele with large effect in KhoeSan population from the far southern Africa. Among the earliest diverged human populations, KhoeSan possess relatively light skin as compared to their Bantu-speaking neighbors. Through our association study in two KhoeSan communities, we found a large effect variant in the canonical pigmentation gene *SLC24A5* that lightens skin pigmentation by 4 melanin units, and explains 12% of the phenotypic variance. This nonsynonymous variant is present at a high allele frequency of 38% in our cohort, which cannot be explained by the proportion of recent gene flow from Europeans. After targeted sequencing this region at high coverage in 430 individuals from this cohort, we demonstrate that the derived haplotypes in KhoeSan are identical to those fixed in Europeans, forming a strong starburst pattern in the network. Time of origin of the derived allele in the two populations are estimated to overlap with each other (18 kya[12 kya – 42 kya] in KhoeSan vs. 13kya[6 kya – 41 kya] in Europeans). The possible introduction of this allele to KhoeSan is about 2~3 kya via gene

flow from eastern African pastoralists, who carried non-African admixture. We test this hypothesis using an approximate Bayesian computation (ABC) approach, incorporating demographic models and selection.
The broader implications and significance include that 1) our finding shows a rare case of strong selection on an allele introduced through introgression in human history; 2) method wise, we demonstrate a novel strategy of tracing the selection on both the genotype and corresponding phenotype, by modeling the signal from the genetic association and its selection through demographic history; 3) as an example of exploring selection in an understudied population, our finding enriches the understanding of the story on convergent evolution of light skin pigmentation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP6

The role of individual transcriptomes in the expression of potentially harmful variation

Isabel Alves [1,*], Hilary Edgington [1], Vanessa Bruat [2], Marie-Julie Favé [2], Philip Awadalla [1 2 3]

[1]Informatics and Bio-computing , Ontario Institute for Cancer Research, Toronto, [2]Department of Pediatrics, Sainte-Justine University Hospital Research Centre, University of Montreal, Montreal, [3]Department of Molecular Genetics, University of Toronto, Toronto, Canada

Abstract: Large-scale population studies have recently shown that most variation in the human genome is rare, functionally relevant and affects phenotypic variation in response to environment. However, it was also shown that individuals carrying rare highly penetrant disease-causing mutations do not manifest their effect. With that being said, what are the mechanisms behind this resilience? For instance, does gene expression play a role in buffering the manifestation of phenotypes associated with deleterious variation?
To address these questions we combined whole-blood transcriptomic and genome-wide SNP-chip data from 650 individuals with European and French-Canadian ancestry (known to harbor larger proportions of damaging functional variants) from the CARTaGENE Project. We estimated allele-specific expression (ASE) proportions per heterozygous site per individual after correcting for "read-mapping bias". We then investigated the allele frequency distribution and characterized the functional relevance of derived mutations experiencing over- or under- allele-specific expression (ASE profile) across individuals.
Our results show that read-mapping bias distorts ASE proportion estimation, as previously proposed, but after correction, fixed or almost fixed mutations tend to be more frequently over-expressed whereas rare derived mutations tend to be more often under-expressed across individuals. Variants that are under-expressed are enriched for potentially deleterious mutations, a pattern not seen among the over-expressed derived mutations.
In sum, we show that ASE affects genes reported to be associated with cardio-metabolic traits and buffers the impact of potentially deleterious mutations on phenotypes, even after stringent corrections. Overall, our results shed light on the influence of individual transcriptomes on disease-related phenotypes and disease susceptibility.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-242

Recombination, evolution and the genetic architecture of fitness-proximal traits in an experimentally evolved C. elegans mapping population

Luke Noble [1,*], Matthew Rockman [1], Henrique Teotonio [2], Boris Shraiman [3]

[1]Biology, New York University, New York, United States, [2]Institut de Biologie, Ecole Normale Superieure, Paris, France, [3]Kavli Institute for Theoretical Physics, University of California, Santa Barbara, United States

Abstract:

Understanding the genetic basis and evolution of complex traits remains a major challenge in biology. The extent of polygenicity, phenotypic plasticity and epistasis are all important for predicting individual phenotypes and for parameterizing models of phenotypic evolution but, for most organisms and traits, their contribution to variation is uncertain.

I will present data from a recombinant inbred line quantitative trait mapping panel we have recently generated for the hermaphroditic nematode Caenorhabditis elegans, which addresses both the genetic architectures of fitness-related traits and how they vary with recombination. The C. elegans multiparental panel of 507 lines was created by hybridization of 16 wild isolates, followed by experimental evolution for >200 generations with the amount of sex genetically controlled, and provides gene-level mapping resolution for alleles of moderate (>5%) effect. Looking, so far, at two challenging, correlated traits - worm fertility and size - we find extreme polygenicity, a paucity of individually significant additive effects, and extensive epistasis. Divergent effects across traits highlights some of the constraints imposed by multivariate selection and linked selection.

I am currently analyzing expression data for these lines and plan also to discuss how genetic control of gene expression, a central intermediary in phenotypic evolution, evolves as a function of recombination.

Understanding the genetic basis and evolution of complex traits remains a major challenge in biology. The extent of polygenicity, phenotypic plasticity and epistasis are all important for predicting individual phenotypes and for parameterizing models of phenotypic evolution but, for most organisms and traits, their contribution to variation is uncertain.

I will present data from a recombinant inbred line quantitative trait mapping panel we have recently generated for the hermaphroditic nematode Caenorhabditis elegans, which addresses both the genetic architectures of fitness-related traits and how they vary with recombination. The C. elegans multiparental panel of 507 lines was created by hybridization of 16 wild isolates, followed by experimental evolution for >200 generations with the amount of sex genetically controlled, and provides gene-level mapping resolution for alleles of moderate (>5%) effect. Looking, so far, at two challenging, correlated traits - worm fertility and size - we find extreme polygenicity, a paucity of individually significant additive effects, and extensive epistasis. Divergent effects across traits highlights some of the constraints imposed by multivariate selection and linked selection.

I am currently analyzing expression data for these lines and plan also to discuss how genetic control of gene expression, a central intermediary in phenotypic evolution, evolves as a function of recombination.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-241

Dissecting the Genetic Basis of Species-specific Essential Genes in Yeast

Monica Sanchez [1,*], Celia Payen [1], Frances Cheong [1], Blake Hovde [1], Jeffrey Skerker [2], Rachel Brem [3], Amy Caudy [4], Maitreya Dunham [1]

[1]Genome Sciences, University of Washington, Seattle, [2]Energy Biosciences Institute, [3]Molecular and Cellular Biology, University of California Berkeley, Berkeley, United States, [4]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada

Abstract: To understand how complex genetic networks perform and regulate diverse cellular processes, the function of each individual component must be defined. One powerful approach to defining gene function is via genetics: phenotypes conferred by mutant alleles can be used to determine what processes depend on the normal function of a gene. This approach has been wildly successful in model organisms, in which comprehensive studies have been performed. These results are often translated to the increasing number of newly sequenced genomes by using sequence homology. However, sequence similarity does not always mean identical function or phenotype, suggesting that new methods are required to functionally annotate newly sequenced species. We have implemented comparative functional analysis by high-throughput experimental testing of gene function in a sequenced, but otherwise understudied species of yeast, *Saccharomyces uvarum*, a sister species of *S. cerevisiae.* We created a haploid and heterozygous diploid library containing Tn7 transposon insertions in *S. uvarum* to identify species-specific essential genes as a first step to identifying genes with divergent function. Using deep sequencing, we identified 50,193 and 37,022 independent insertion sites found in 80% and 63% of orthologous coding sequences in the diploid and haploid libraries respectively. We predict 809 genes to be essential in *S. uvarum*, with 444 of those genes also known to be essential in *S. cerevisiae*. We predict 124 genes to be *S. cerevisiae*-specific essential genes (non-essential in *S. uvarum),* and 156 genes to be *S. uvarum*-specific essential genes. We created 12 heterozygous deletion mutants to validate highly ranked genes in species-specific essential and non-essential gene categories. However, these genes are capable of cross-species complementation, demonstrating that differences in genetic background must contribute to differential gene essentiality. We are now mapping these genetic background elements to determine the molecular explanation for these differences. This data set provides direct experimental evidence of gene function across species, which can inform comparative genomic analyses, improve gene annotation and be applied across a diverse set of microorganisms to further our understanding of gene function evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-412

POST-TRANSCRIPTIONAL REGULATION AFFECTS DEVELOPMENT OF DROSOPHILA PIGMENTATION, A MODEL SYSTEM FOR STUDYING PHENOTYPIC EVOLUTION

Abigail Lamb [1,*], Patricia Wittkopp [2]

[1]Molecular, Cellular, and Developmental Biology, [2]Ecology and Evolutionary Biology; Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI, United States

Poster: Studies identifying the genetic basis of phenotypic changes frequently support the hypothesis that morphological evolution is more likely to proceed through non-coding mutations than through changes in the coding sequence of genes. Much of the theoretical framework underlying this “cis-regulatory hypothesis” is focused on mutations in cis-regulatory DNA which lead to spatiotemporally restricted changes in mRNA expression. However, post-transcriptional regulation can also prevent mRNA from being translated to protein in a spatially and/or temporally restricted manner. Furthermore, a large proportion of studies used to identify the genetic mechanisms underlying phenotypic evolution rely on pre-existing knowledge of genetic networks to identify candidate genes, but few model systems have well-characterized post-transcriptional regulatory networks. The biosynthetic pathway underlying *Drosophila* pigmentation is well characterized, and many potential and confirmed transcription factor regulators of melanin synthesis genes have been identified, making it an excellent model of gene regulatory evolution. But little is known about the role of post-transcriptional regulation in the development of *Drosophila* pigmentation. Here we present the results of a genetic screen to identify microRNAs that affect pigmentation in *D. melanogaster* in order to expand this important model system to allow the inclusion of post-transcriptional regulation in future evolutionary studies.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP1

Deep mutational scanning of an ancient protein: the many roads not taken in protein evolution

Joseph Thornton*

Abstract: To understand why evolution unfolded as it did, we must reconstruct not only the history that took place but also the many possible histories that did not. We combined ancestral protein reconstruction with deep mutational scanning to comprehnsively characterize the sequence space around an ancient protein that evolved a biologically essential new function in the deep past -- a transcription factor that acquired the capacity to recognize a new DNA target during early vertebrate evolution. This strategy allowed uys to compare the pathway that the protein followed during its actual history to a vast ensemble of alternative evolutionary outcomes and trajectories. We found hundreds of different genetic outcomes and mutational pathways by which the ancestral protein could have evolved the new function; many of these are functionally superior and easier to access from the ancestral starting point than the historical outcome was, and they use different biochemical mechanisms to recognize DNA. All pathways to the new function would have required permissive epistatic substitutions, but in some cases the permissive steps are different from those that occurred during history. If the protein's evolution had begun from slightly different ancestral starting points, different ways of achieving the new function would have become accessible, some of which would have required no permissive substitutions at all. By mapping the historically relevant portion of a protein's vast sequence space, these results reveal the wild contingency and stochasticity of molecular evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-238

Laboratory Evolution of E. coli to Variable Substrate Environments Yields Distinct Phenotypic and Genetic Adaptive Strategies

Troy Sandberg*, Colton Lloyd, Bernhard Palsson, Adam Feist

Abstract: Adaptive Laboratory Evolution (ALE) experiments are often designed to maintain a static culturing environment to minimize confounding variables that could influence the adaptive process, but dynamic nutrient conditions occur frequently in both natural and bioprocessing settings. To study the nature of carbon substrate fitness tradeoffs, we evolved batch cultures of *E. coli* via serial propagation into flasks alternating between glucose and either xylose, glycerol, or acetate. Genome sequencing of evolved cultures revealed several genetic regions preferentially mutated under dynamic conditions, and different adaptation strategies depending on the substrates being switched between – in some environments a persistent "generalist" strain developed while in another, two "specialist" subpopulations arose that alternated dominance. Diauxic lag phenotype varied across the generalists and specialists, in one case being completely abolished, while gene expression data distinguished the transcriptional rearrangements implemented by strains in pursuit of growth optimality. Genome-scale metabolic modeling techniques were then used to help explain the inherent substrate differences giving rise to the observed distinct adaptive strategies. This study gives insight into the population dynamics of adaptation in a highly variable environment, as well as the underlying metabolic and genetic mechanisms. Furthermore, ALE-generated optimized strains have phenotypes with potential industrial bioprocessing applications.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP14

Transcriptomic Architecture Influences Phenotypic Diversification across Vertebrates

Rebecca Young [1,*], Heather Goldsby [2], Arend Hintze [3], Hans Hofmann [1]

[1]Integrative Biology, University of Texas, Austin, [2]Department of Computer Science, [3]Department of Integrative Biology, Michigan State University, East Lansing, United States

Abstract: Despite the astonishing diversity of animal phenotypes the molecular mechanisms underlying evolutionary diversification are remarkably conserved. In the early 19th century, von Baer noticed that embryos of a given lineage exhibit greatest morphological similarity across species during mid-embryogenesis (a developmental pattern resembling an hourglass), a phenomenon that since has been referred to as the 'phylotypic period.' This hourglass model is supported by recent comparative transcriptomic analyses, possibly due to interdependence (pleiotropy, epistasis) of signaling networks during the phylotypic period, as multiple organ systems develop throughout the embryo. We hypothesized that the topology of embryonic co-expression networks should also transition from more modular before and after the phylotypic period to highly interconnected during the phylotypic period. We first characterized phenotypic variation in embryogenesis across species. We then conducted a meta-analysis of gene co-expression across developmental stages in five vertebrate species and examined the developmental trajectory of transcriptomic networks across species. Finally, we use a novel *in silico* evolutionary model of a developing tissue to test different mechanistic hypotheses of phenotypic diversification. Our results demonstrate how transcriptomic architecture is associated with the generation of phenotypic variation across species. This research reveals new complex relationships in the evolution of phenotypic diversity and its molecular underpinnings.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-237

Does genetic correlation constrain or facilitate long-term phenotypic evolution?

Wei-Chin Ho [1,*], Jianzhi Zhang [1]

[1]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, United States

Abstract: The evolution of a phenotypic trait can be influenced by its genetic correlations with other traits. While genetic correlations may constrain phenotypic evolution if antagonistic pleiotropy is prevalent, it can also facilitate phenotypic evolution if pleiotropy is mostly concordant. Past studies on these possibilities focused on short-term adaptations on the order of 1-10 generations, leaving the impact of genetic correlations on long-term evolution unclear. Here we address this question using two large sets of traits in yeast: 210 morphological traits and 3466 gene expression traits, which have been shown to be respectively under generally adaptive and neutral evolution among natural yeast strains that have diverged from one another for 10^6 to 10^7 generations. For both morphological and gene expression traits, we found no significant correlation between the among-strain phenotypic difference for a trait and the average genetic correlation between the trait and all other traits. Therefore, at least for the two distinct types of traits considered, genetic correlations neither facilitate nor constrain long-term phenotypic evolution in general. Because the datasets used are large and the phenotypic measurements are quite accurate, our finding is unlikely due to a lack of statistical power but is likely to be genuine.

Expanded summary*: While it has been a long time people study whether genetic correlation constrain or facilitate phenotypic evolution, there are some limits in the previous literatures. First, because of researchers' interests, people tend to focus on the traits under directional selection. Second, previous studies only focus on short term evolution (~10 generations) and ignore the overall trend in long-term evolution. Therefore, we focus on two sets of traits: 210 morphological traits and 3466 gene expression levels, which have been shown generally under adaptive evolution and neutral evolution, respectively. In addition, the phenotypic evolution data among different strains are available, which could represent a long-term evolution outcome.

For the morphological traits, we took advantage a data set of 210 traits measured in 4718 single-gene deletion lines and 37 yeast strains. We first calculated the raw effect size (ES_{ij}) of deleting gene *i* on trait *j*, which equals to $|x_{ij} - w_j|/w_j$, where x_{ij} is the mean phenotypic value of trait *j* in the deletion strain *i*, and w_j is the mean phenotypic value of trait *j* in the BY wild-type. To reduce the measurement noise, we calculated net *ES* subtract each *ES* by the pseudo *ES* based on 1000 random sample of absolute phenotypic differences between different BY replicates. After acquiring these net *ES*, we calculated the genetic correlation between traits *j* and *k* by Pearson correlation coefficient between two vectors (net ES_{1j}, net ES_{2j}, …, net $ES_{4718,j}$) and (net ES_{1k}, net ES_{2k}, …, net $ES_{4718,k}$). Similarly, for yeast strain *i* and *j*, we first calculated the raw evolutionary distance for the trait *j* between strains *p* and *q* by $ED_{jpq} = |x_{jp} - x_{jq}| / [(x_{jp} + x_{jq})/2]$, where x_{jp} and x_{jq} are the mean phenotypic values of a trait *j* in strains *p* and *q*, respectively. We also calculated pseudo *ED* by 100 sets of bootstrapped samples using same strain, and the net *ED* equals raw *ED* subtracting pseudo *ED*.

When testing our question, for each trait, we calculated mean genetic correlation (GC), which is the average of genetic correlation between the focal trait and each of other 209 traits. We also calculated mean net ED among all net *ED* of each pair of strains. If GC generally constrains the ED, it is expected to see a negative correlation between them; if GC generally facilitates the ED, it is expected to see a positive correlation between them. As a result, we found the Spearman's correlation coefficient (ρ) between mean GC and mean ED is -0.069 (p-value = 0.32). This no-correlation is not merely a byproduct of various mutational sizes (MS) among different traits, given that the correlation between GC and ED/MS is still not significant (ρ=0.0034; p-value = 0.96), where MS is measured by mean of net *ES* for each trait. These results suggest GC does not neither constrain nor facilitate the phenotypic evolution.

To make sure our methodology does not artificially cause any positive or negative correlations, we randomly shuffle the matrix of net *ES* and perform similar analysis using the random shuffled *ES*. The mean and median of these Spearman's correlation coefficients (ρ) between mean GC and mean ED is -0.017 and -0.016, respectively, and the standard deviation is 0.10. Given that there are 30% of these randomized ρ's smaller than -0.069, it is likely genuine that GC and ED has no significant correlation.

We also performed similar analysis to 3466 gene expression traits, where ES is measured in 1486 single-gene deletion lines. ED is measured between two yeast strains BY and RM. We found that the Spearman's correlation coefficient between mean GC and mean ED is 0.025 (p-value = 0.20). Therefore even in traits generally undergone neutral evolution, GC still does not neither constrain nor facilitate the phenotypic evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-240

Underlying mechanisms of evolvability in experimentally evolved populations

Rachel Staples*, Kelly Phillips [1], Timothy Cooper [1]

[1]University of Houston, Houston, United States

Abstract: Genetic background shaped by prior adaptation may influence evolvability, or the ability for future adaptive evolution, particularly in the face of environmental change. We previously found that experimentally evolved *Escherichia coli* populations that had adapted to a glucose environment were more evolvable than their naive ancestor in a novel lactose environment. We theorize that this increase in evolvability is the result of new compensatory mutations acquired during lactose adaptation that neutralized the negative fitness effects of previously-substituted, now deleterious alleles. This compensation effectively unmasked the underlying potential of existing beneficial alleles, quickly increasing fitness. Here we provide evidence that replacing candidate alleles in both evolved and naive genetic backgrounds significantly influences fitness, suggesting that rapid adaptation through compensation may be a mechanism of evolvability.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-239

The genetic evolution of beak shape in Birds

Joseph Palmer [1,*], Chris Cooney [1], Toni Gossmann [1]

[1]University of Sheffield, Sheffield, United Kingdom

Abstract: A central problem in the genomic era concerns the genetic underpinning of phenotype. Birds have long been the focus of pioneering research in evolution and have helped to answer some of the most important questions in evolutionary biology. Beak shape is emblematic of evolution and given the wealth of genetic data produced by projects such as Bird 10K, studies examining the genetic basis of these complex and highly ecologically-relevant appendages have made exiting inroads. This study builds on recent findings and seeks to extend our understanding of avian beak genetics by examining over 60 fully sequenced bird genomes. This data is then mapped to morphological changes in beak shape identified by research on over 2000 species, representing over 97% of extant genera. A major question at the heart of this research concerns the relative roles of coding and regulatory genetic changes in explaining phenotype evolution. Current evidence for coding sequence changes as a determinant of phenotypic evolution far outweighs that of regulatory changes. However, regulatory changes may be more subtle due to their pleiotrophic nature. We identify sequence changes in coding and regulatory regions associated with beak shape morphology in a phylogenetic context and determine their fitness effects both within and between species. Our research unites comprehensive genetic data with the most advanced morphometric measurements of beak phenotype to map the genetic evolution of one of the most ecologically important traits in the animal kingdom.

Statement: I am a Masters Of Research student investigating avian genetics at the University of Sheffield. My research has exposed me to a wealth of literature and in taking the early steps in my career, I am building on gaps in our understanding . However, absorbing literature represents only part of what it takes to produce excellent research. Talking to other researchers is vital, both for identifying new developments in the field and in facilitating future collaborations. Whilst I regularly attend meetings with university research groups, my interaction with the wider scientific community comes only from attending talks from visiting researchers. As yet, I have not had the opportunity to attend a scientific conference. My research has familiarised me with key scientists in my field and the opportunity to hear about and discuss their research in more depth would be extremely valuable. Most notably, the symposiums on the Mechanisms of Phenotypic Evolution, Mutational Mechanisms, Evolution of Gene Regulation, and the Evolution of Complex Traits are highly relevant to my interests and would provide unparalleled insight into furthering my research. Having a mentor to my first conference will greatly enhance my experience and I relish the opportunity to present some of my own work. Beyond interacting with established figures in the field, I am also keen to engage with other early career researchers and share experiences. Taking my first steps to becoming a researcher, I feel this opportunity would improve my research and facilitate my integration within the global scientific community.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-236

Regulatory control of phenotypic plasticity - Dissecting the molecular basis of inducible defences in Daphnia

Dörthe Becker [1,*], Andrew P. Beckerman [2]

[1]Department of Biology, University of Virginia, Charlottesville, United States, [2]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom

Abstract: Even though it is generally acknowledged that phenotypic plasticity is an important property of developmental systems which allows organisms to cope with environmental unpredictability and/or heterogeneity, its role in adaptive evolution remains contentious. Empirical data suggests that phenotypic plasticity can increase organism survival under specific conditions. However, our understanding of the cellular processes by which plastic responses to environmental stimuli are triggered remains unresolved. Revealing such mechanism is important to understand how plasticity influences diversification and the rate of evolution.
We investigated the mechanistic basis of predator-induced plasticity in the freshwater invertebrate *Daphnia*, an ecological and genetic model system. Invertebrate midge predation triggers a well-established example of plasticity, and induced morphological change called neckteeth. These neckteeth are providing a clear fitness benefit in terms of reduced predation. The underlying regulatory mechanisms that form the basis of this trait are still unknown.
We combined phenotypic, metabolomics and genomic screening of distinct *D. pulex* genotypes with highly contrasting levels of induced defence. We (1) profiled the morphology and associated life-history changes; (2) examined the gene expression profile and resulting gene regulatory networks of daphnids exposed to predation risk; and (3) identified unique metabolomic signatures that reflect the functional, phenotypic response of daphnids to the specified gradient of environmental risk(s).
Bridging the gap between genotype and phenotype remains a major focus in evolutionary research. Using a textbook example of plasticity, our work allows a rare glimpse into how two molecular mechanisms combine to drive variation in phenotypic plasticity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-256

Exploring the many facets of phenotypic changes after 13 years of divergent selection for flowering time in maize

Maud Tenaillon [1,*], Adrienne Ressayre [2], Aurélie Bourgais [2], Hélène Corti [2], Martine Le Guilloux [1], Elodie Marchadier [3], Christine Dillmann [3]

[1]Génétique Quantitative et Evolution - Le Moulon, CNRS, Université Paris-Saclay, [2]Génétique Quantitative et Evolution - Le Moulon, INRA, Université Paris-Saclay, [3]Génétique Quantitative et Evolution - Le Moulon, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France

Abstract: Two seed lots from maize inbred lines were chosen as initial populations for a divergent selection experiment for flowering time. After 13 generations within each line, we obtained two populations Late and Early – further structured into families – that display major differences in flowering time. We observed a significant response to selection in most families, indicating a significant heritability for different traits related to flowering time despite strong genetic drift. We aimed at exploring convergence (within Early/Late) and divergence (between Early and Late) among genotypes. We first analysed the transcriptomes of the meristems of evolved individuals during flowering transition to identify genes differentially expressed. Second, we characterized plants growth and development in the field. RNA sequencing revealed about hundred genes differentially expressed between genotypes during floral transition. Examination of phenotypic changes suggest that selection has triggered modifications of the timing of the developmental program (transitions) rather than of the developmental rates. Altogether, our results indicate that phenotypic divergence and convergence are achieved through distinct developmental “routes” in the different families. Other facets of phenotypic changes will be discussed.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-252

The Satellite MC1R allele associated with light plumage color in ruffs (Philomachus pugnax) involves both regulatory and structural changes.

Doreen Schwochow-Thalmann [1 2,*], Ilona Mandrika [3], Ance Roga [3], Davids Fridmanis [3], David B. Lank [4], Leif Andersson [5 6]

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, [2]Génetique Animale et Biologie Intégrative, INRA/ AgroParisTech, Paris, France, [3]Latvian Biomedical Research and Study Center, Riga, Latvia, [4]Department of Biological Sciences, Simon Fraser University, Burnaby, Canada, [5]Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden, [6]Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences , Texas A&M University, College Station, United States

Abstract: The ruff is a bird with a complex mating system and unusual plumage diversity. Males belong to one of three alternative reproductive strategies: 'independents' occupy and aggressively defend territories on leks, 'satellites' are non-territorial and submissive, and 'faeders' are female mimics. During breeding season, independents and satellites grow elaborate ornamental feathers around the neck and head, whereas faeders maintain non-ornamental, female-like plumage all year around. Ornamental plumage pigmentation is most variable in independent males, which range from white to browns and black with various forms of black patterning. Satellites have predominantly white feathers, with variation in brown and almost no black patterning or feathers. A 4.5 Mb inversion has recently been identified as the underlying cause for the alternative male morphs. The melanocortin receptor 1 gene (MC1R) is located within the inverted region and four amino acid substitutions of conserved residues were identified on the derived satellite allele. We have investigated the gene expression levels of MC1R in growing feather follicles of ornamental and non-ornamental feathers in independents and satellites. We also performed allelic imbalance assays of differently pigmented satellite feathers and *in-vitro* experiments to measure MC1R signaling in a cell culture model. The results suggest that combinations of regulatory changes are associated with white ornamental feathers in satellites.

Expanded summary*: My research focuses on developing a better understanding of the genetic basis underlying specific coloration patterns in birds. The ruff (*Philomachus pugnax*), a wading bird with a very impressive mating system and plumage variation, is one of my focal species. We were able to identify the genetic basis of three different male reproductive strategies in this bird: the independent, the satellite and the faeder – which intriguingly correlate with pigment intensity of their respective breeding plumage. Our current results indicate that variation at the melanocortin receptor 1 gene located within the inversion is playing an important role in the mechanism controlling this variation.

Identifying new genes involved in avian pigmentation can be challenging in wild birds, which is why I am using the chicken as a model system. Humans have bred individuals for a certain appearance ultimately resulting in the astonishing color variation observed in modern chicken breeds. I have generated NGS data to explore the causative mutation of *Patterning*, one of the major drivers of pigment patterns on individual chicken feathers. In a next step I will investigate how the genetic variant alternate the function of the involved protein. In the case of sex-linked barring - a plumage pattern in which pigmented and apigmented bars are alternating on an individual feather - I was able to show that a combination of regulatory and missense mutations in *CDKN2A* is creating this unique pattern. Sex-linked barring most likely resulted from evolution of alleles: first a regulatory mutation increased *CDKN2A* expression in the growing feather causing a basic, more dilute barring pattern, secondly two different missense mutations impairing protein function, refined and intensified the black bar. Intriguingly, although the *CDKN2A* is a tumor suppressor gene and involved in familial melanoma in humans, sex-linked barred chickens do not show a higher prevalence for tumor development although they are expressing a malfunctioning protein.

Conclusively, my research is contributing to a more comprehensive understanding of the genes and pathways involved in pigmentation, which aside from its evolutionary merits might also have implications for human diseases. My work will shed light on

how complex traits in general do evolve over time and how mutations shape a phenotype by interactions with the environment, creating i.e. sophisticated camouflage abilities.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OM-OS7

Long-term patterns of bacterial phenotypic divergence

German Plata [1,*], Dennis Vitkup [1]

[1]Department of Systems Biology, Columbia University, New York, United States

Abstract: Understanding the breadth of prokaryotic diversity and the evolutionary processes that shape it entails an investigation of microbial phenotypic properties. It is phenotypes rather than gene sequences what defines the activities, interactions, environmental roles, and evolutionary fates of microorganisms; however, a thorough examination of phenotypic evolution is complicated due to a paucity of experimental data. The use of *in silico* models of bacterial metabolic networks allows the prediction of phenotypes from their annotated genomes. We have applied constraints based analyses of metabolic networks to predict and compare phenotypic profiles of hundreds of bacterial species, and to observe patterns of phenotypic similarity over times that span a few million to billions of years of evolutionary divergence. Our results demonstrate that while on short timescales phenotypic similarity shows a low correlation with genetic distance, it displays a remarkably consistent exponential decay over long evolutionary periods. Similar divergence patterns but different evolutionary rates and conservation levels are observed for distinct phenotypes, including the ability to use different nutrients or the effects of gene deletions. These results are consistent with experimental data gathered from independent sets of phylogenetically diverse species. The observed divergence patterns, with a constant fraction of phenotypic changes per unit time, are reminiscent of the molecular clock in protein evolution. Similar to molecular evolution, the evolution of microbial phenotypes can be understood in the context of an evolving fitness landscape, pointing at the processes by which prokaryotes conquered Earth's environments.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-258

Dissecting the function and evolution of complex gene interaction networks with haploid genetics in Nasonia wasps

Jeremy Lynch [1,*], Lorna Cohen [1], John Werren [2]

[1]Biological Sciences, University of Illinois at Chicago, Chicago, [2]Biology, University of Rochester, Rochester, United States

Abstract: It is increasingly clear that the evolution of developmental processes often involves changes within complex networks of interacting genes. The identity of the participants and the nature of the interactions within the networks are usually obscure, but can be revealed by the phenomenon of epistasis, where the novel combination of alleles leads to a phenotype significantly different from the sum of the phenotypes of the alleles in isolation. Recently, genome sequencing and other powerful tools, which can be used to explore the complex genetic basis of epistasis, have become widely available. However, studying epistasis is still difficult in typical diploid animal model systems, due to dominance interactions between alleles within loci, and the rapid increase in rarity of desired genotypes as the number of interacting loci increases. *Nasonia* wasps can address these difficulties, as all males are haploid, thus eliminating dominance interactions and increasing the frequency of desired genotypes. Viable and fertile hybrids between *Nasonia* species with morphologically distinct males can be made, and recombinant haploid F2 males are readily obtainable. With genome sequences available for the relevant species, these features make *Nasonia* a powerful system for evolutionary genetics. We will present our progress in identifying the genetic and developmental basis of two phenomena that arise through complex networks of interacting genes. One is the distinctive "cheeks" observed in *Nasonia giraulti* males, but not in any other *Nasonia* species. The other is the facial clefting observed only in F2 haploid hybrid males between *N. giraulti* and *N. vitripennis.*

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP12

Gene flow, ancient polymorphism and ecological adaptation shape the genomic landscape of divergence among Darwin's finches

Fan Han [1,*], Sangeet Lamichhaney [1], B. Rosemary Grant [2], Peter Grant [2], Leif Andersson [1 3 4], Matthew Webster [1]

[1]Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden, [2]Ecology and Evolutionary Biology, Princeton University, New Jersey, United States, [3]Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, [4]Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas, United States

Abstract: Genomic comparisons of closely related species have identified "islands" of locally elevated sequence divergence. Genomic islands may contain functional variants involved in local adaptation or reproductive isolation, and may therefore play an important role in the speciation process. However, genomic islands can also arise through evolutionary processes unrelated to speciation and examination of their properties can illuminate how new species evolve. Here, we performed scans for regions of high relative divergence (F_{ST}) in 12 species pairs of Darwin's finches at different genetic distances. In each pair, we identify genomic islands that are on average elevated in both relative divergence (F_{ST}) and absolute divergence (d_{XY}). This signal indicates that haplotypes within these genomic regions became isolated from each other earlier than the rest of the genome. Interestingly, similar numbers of genomic islands of elevated d_{XY} are observed in sympatric and allopatric species pairs, suggesting that recent gene flow is not a major factor in their formation. We find that two of the most pronounced genomic islands contain the *ALX1* and *HMGA2* loci, which are associated with variation in beak shape and size, respectively, suggesting that they are involved in ecological adaptation. A subset of genomic island regions, including these loci, appears to represent anciently diverged haplotypes that evolved early during the radiation of Darwin's finches. Comparative genomics data indicate that these loci, and genomic islands in general, have exceptionally low recombination rates, which may play a role in their establishment.

Expanded summary*: Genomic islands of divergence, identified by comparing closely related species, can be generated from different evolutionary processes. Some gnomic islands that contain variation associated with reproductive isolations are related to speciation, and the others caused by ancient polymorphisms, recurrent selection or other processes are not. Identifying genomic islands that are formed by different mechanisms can uncover the evolutionary history of species, and therefore understand the genetic background underlying speciation. Darwin's finches are a classic model for study of speciation. The recognized 18 species evolved from a common ancestor and rapidly adapted to different ecological conditions. They evolved within an isolated environment and diversified in beak morphology and feeding habits. From 12 pairwise comparisons of species at different genetic distances, we showed the genomic landscape of divergence among Darwin's finches. The genomic islands identified using F_{ST} statistics display several distinguished features. First, they are elevated in absolute divergence d_{XY} in addition to F_{ST} in most of sympatric and allopatric comparisons. This indicates the divergence in the island regions started to form earlier than the radiation of the finches, and recent gene flow is not a major cause of the islands. Second, *ALX1* and *HMGA2*, which are associated with beak morphologies, locate in two of the most pronounced genomic islands. This suggests that a subset of the genomic islands are involved in ecological adaptation. The deep divergence of the two haplogroups of *ALX1* and *HMGA2* furthermore implies some of the genomic islands evolved from ancient polymorphism. Third, we found the genomic islands in general have significantly low recombination rates than the genomic average, suggesting the role of background selection played in the establishment of the genomic islands. Our data are consistent with an evolutionary scenario where genomic islands of divergence have evolved throughout the radiation of the Darwin's finches, most often in regions with low recombination and often due to genetic adaptation, as exemplified by the beak loci.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP2

Distinguishing between mechanistic tradeoffs and mutational entropy in protein evolution

Christopher S. Wylie [1], Caleb Weinreb [2], Daniel Weinreich [3,*]

[1]Insight Data Science, Providence, RI, [2]Systems Biology, Harvard Medical School, Boston, MA, [3]Ecology and Evolutionary Biology, Brown University, Providence, RI, United States

Abstract: The mutational input that fuels evolution may already have statistical pleiotropic biases, which then might be manifest in phenotypic correlations among extant organisms. Using TEM-1 ß-lactamase as an example, we provide an experimental framework for distinguishing between such entropic biases in observed phenotypic correlations, and the more common explanation: mechanistic tradeoffs embedded in the biochemistry and biophysics of the protein.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-255

Using position effect variegation in Drosophila melanogaster eyes to investigate cryptic genetic variation

Joyce Kao [1,*], Daniel McNelis [1], Andrew Tan [1], Mark Siegal [1]

[1]Biology, New York University, New York, United States

Abstract: Organisms constantly adapt and evolve in response to internal and external stimuli. Changes can arise by either new mutations or by pre-existing genetic variation in the system. The latter standing genetic variation, which does not affect the usual range of phenotypes, but can potentially modify a phenotype in the event of an environmental or genetic perturbation, is cryptic genetic variation (CGV). To demonstrate the release of CGV, we perturb the gene, *His2Av*, in *Drosophila melanogaster* (fruit fly). *His2Av* encodes an alternate histone and is the ortholog of yeast *HTZ1*, which contributes to robustness of cell morphology against microenvironmental variation in yeast. We created a new GFP-tagged *His2Av* knockout mutation and introgressed it via backcrossing into the genetic backgrounds of the Drosophila Genetic Reference Panel (DGRP), which is a panel of naturally-derived inbred fly lines. We reveal dominant effects of the *His2Av* knockout in different DGRP genetic backgrounds by crossing our mutation-introgressed lines to the w[m4] line, which is a fly line where an inversion on the X-chromosome places the *white* gene encoding red pigment in eyes next to pericentric heterochromatin. We developed a new image-based position effect variegation assay to quantify pigment variegation in fly eyes. Quantifying the amount of variation in phenotypic responses within and between DGRP lines and with and without *His2Av* will help elucidate the prevalence and behavior of CGV.

Expanded summary*: Organisms constantly adapt and evolve in response to internal and external stimuli. Changes can arise by either new mutations or by pre-existing genetic variation in the system. The latter standing genetic variation, which does not affect the usual range of phenotypes, but can potentially modify a phenotype in the event of an environmental or genetic perturbation, is cryptic genetic variation (CGV). Studying cryptic genetic variation in present times is especially important. Due to the environmental and cultural changes in recent human history, the uncovering of CGV could be a potential explanation for the rising incidences of "diseases of modernity" such as asthma, diabetes, depression, etc. (Gibson, 2009). It has also been suggested that CGV can impact adaptation to climate change (Gibson & Reed 2008; Berger *et al.*, 2011).

We believe that cryptic variants are quite prevalent in the genome and do play a significant part in determining the consequences of genetic perturbations in an organism. To demonstrate the release of CGV, we use the fruit fly, *Drosophila melanogaster* and perturb the gene, *His2Av*, which encodes an alternate histone. Work done in nematode worms and yeast reveal that chromatin regulation may be a factor in suppressing cryptic variation (Lehner *et al.*, 2006; Tirosh *et al.*, 2010). Additionally, *His2Av* is the ortholog of yeast *HTZ1*, which contributes to robustness of cell morphology against microenvironmental variation (Levy & Siegal, 2008).

We have created a new GFP-tagged *His2Av* knockout mutation and introgressed this mutation via backcrossing into the genetic backgrounds of the Drosophila Genetic Reference Panel (DGRP), which is a panel of naturally-derived inbred *D. melanogaster* lines. We then crossed each of our introgressed lines to the w[m4] line to reveal dominant effects of the *His2Av* knockout in different DGRP genetic backgrounds. The w[m4] line has an inversion on the X chromosome that places the *white* gene encoding red pigment in eyes next to pericentric heterochromatin so that chromatin state determines whether or not *white* is expressed thereby creating mottled eye pigmentation. We developed a new image-based position effect variegation assay to quantify pigment variegation in fly eyes. By quantifying the amount of variation in perturbed phenotypic responses within and between DGRP lines, and with and without the *His2Av* mutation, we can understand the prevalence and behavior of cryptic genetic variation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-243

Detecting balancing selection as a maintenance mechanism of psychiatric disorders within human populations.

Daiki Sato [1,*], Takashi Makino [1], Masakado Kawata [1]

[1]Department of Ecology and Evolutionary Biology, Graduate School of Life Sciences, Tohoku University, Sendai, Japan

Abstract: Recently, growing attention has been paid to evolutionary psychiatry, where the evolutionary mechanisms behind the prevalence of psychiatric disorders have been discussed. Though mental health problem has become very common in our modern society, it is hard to explain this phenomenon because maladaptive features should have been removed from population over the course of evolution. To answer this question, we tested a hypothesis that psychiatric disorders-risk alleles have been maintained within populations by natural selection. Out of 1,013 psychiatric disorders-relevant genes, four genes, *CLSTN2*, *FAT2*, *VMAT1* and *ZBBX*, have been detected as positively selected genes in human lineage. Among them, *VMAT1*, vesicular monoamine transporter 1, was particularly interesting since it has a human-unique polymorphism (Thr/Ile) in the 136th amino acid site compared to other mammals (Asn). In addition, 136Ile has been proposed to be linked to bipolar disorder and anxiety although the allele could be found at relatively high frequencies (20-40%) in non-African populations. In non-African populations, Tajima's *D* value showed a significant peak centering around Thr136Ile, suggesting that this polymorphism has been maintained by balancing selection. Our study demonstrates an evolutionary mechanism behind psychiatric disorders and provides insights into the evolution of diseases.

Expanded summary*: Psychiatric disorders, such as schizophrenia, depression and autism, can be seen at relatively high incidence (1-20%) all over the world. It has been known that genetic components greatly contribute to their onset, and that those disorders do degrade patients' quality of life. However, why and how the causative genes have been maintained within current human populations still remains to be elucidated. Revealing its mechanisms is of particular importance not only from the evolutionary perspective but for medical treatment of those disorders. In regard to this question, we focused on a human-unique polymorphism (Thr/Ile) in the 136th amino acid site of *VMAT1* gene, vesicular monoamine transporter 1, detected by the sequence comparison using mammal genome data. This gene encodes the protein importing monoamine, such as serotonin or dopamine, into presynaptic vesicles and mediates neurotransmission. 136Ile has been proposed to be linked to psychiatric disorders such as bipolar disorder or anxiety, but 1000 Genomes Data showed that 136Ile exists at relatively high frequencies (20-40%) in all but African populations. The results of Tajima's *D* values indicated that the allelic polymorphism (136Thr/Ile) has been maintained by balancing selection, such as negative frequency-dependent selection, heterozygote advantage or spatially variable finesses. Moreover, the analysis using coalescent simulations revealed that 136Ile originated around 100,000 years ago, generally regarded as the timing of Out of Africa event by modern humans, giving us insight on the relationship between Out of Africa event and the prevalence of psychiatric disorders in modern humans. Our results suggest that polymorphic alleles causing psychiatric disorders could have been maintained by balancing selection, and that a genetic change coinciding with the global spread of modern humans might have contributed to the risk prevalence of those disorders.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP3

Complex evolutionary roots of skin-barrier function in humans

Eaaswarkhanth Muthukrishnan [1 2,*], Duo Xu [2], Colin Flanagan [2], Margarita Rzhetskaya [3], M. Geoffrey Hayes [3], Ran Blekhman [4], Nina G. Jablonski [5], Omer Gokcumen [2]

[1]Population Genomics and Genetic Epidemiology Unit, Dasman Diabetes Institute, Kuwait City, Kuwait, [2]Department of Biological Sciences, University at Buffalo, The State University of New York at Buffalo, Buffalo, NY, [3]Division of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, [4]Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN, [5]Department of Anthropology, Pennsylvania State University, University Park, PA, United States

Abstract: Rapid evolution of human skin has left evolutionary signatures in the genome. The filaggrin (*FLG*) is a well-studied gene essential for natural skin-barrier function in humans. The extensive genetic variation in this gene, especially common loss-of-function (LoF) mutations, has been strongly associated with atopic dermatitis susceptibility in multiple human populations. To investigate the evolution of this gene, we analyzed 2,504 human genomes and genotyped the copy number variation of filaggrin repeats within *FLG* in 126 individuals from diverse ancestral backgrounds. We were unable to replicate a recent study claiming that LoF of *FLG* is adaptive in northern latitudes with lower ultraviolet light exposure. Further, we integrated results from evolutionary and population genetic analyses, RNA expression data from multiple tissues, as well as molecular mechanisms of *FLG* function to reach multiple striking conclusions: (i) *FLG* LoF mutations are unusually common when compared to other genes in the genome, (ii) the most parsimonious explanation for this observation is that *FLG* LoF variants have no obvious fitness effects despite predisposing to skin disease, (iii) some of the atopic dermatitis susceptibility variants "hitchhike" a selective sweep involving neighboring hornerin (*HRNR*) gene, and (iv) the haplotype block that is under selection carry multiple functional variants, including those that change the expression level of multiple transcripts and microbiome diversity on the skin. Our results present multiple lines of evidence that involve complex evolutionary trajectories in shaping the skin-barrier function in humans.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-245

Assessing the evolution of adaptive coloration using population genetic tools

Kristen Irwin*, Susanne Pfeifer, Stefan Laurent, Jeffrey Jensen

Abstract: Animal coloration has been linked to fitness through many avenues, including crypsis, aposematism, mimicry, thermoregulation, and sexual selection. More recently, evidence for selection on coloration traits has been gathered using population genetic inference. We here review these searches, including the role of selection in each type of adaptive coloration, the adaptive walks characterizing them, and any common genomic signatures. Further, we discuss demographic events that may resemble patterns of selection at the genomic level, and give guidance for parsing the two. We draw attention to recent analyses in lizards, mice, and lepidopterans that not only showcase best practices for accurately identifying genetic regions under selection, but also indicate the various modes of selection that operate on coloration.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-244

A key role for an omega-3 fatty acid desaturase in stickleback freshwater colonization

Jun Kitano [1,*], Asano Ishikawa [1]

[1]Population Genetics, National Institute of Genetics, Mishima, Japan

Abstract: The colonization of empty niches can trigger adaptive radiation, the evolution of ecological and phenotypic diversity within a rapidly multiplying lineage. However, not all lineages have taken advantage of these ecological opportunities. What genetic factors underlie the differences in the ability to colonize new niches? One of the remarkable examples of adaptive radiation is the diversification of threespine sticklebacks: during the last glacial cycles, many freshwater lakes and rivers were newly formed, and marine ancestral sticklebacks have colonized these empty niches and diversified. However, closely related Japan Sea sticklebacks could not colonize freshwater. Our genomic analysis as well as genetic manipulation experiment showed that a “copy & paste” gene transposition of an omega-3 fatty acid desaturase gene occurred in the threespine stickleback lineage, but not in the Japan Sea lineage, which could explain some of the differences in the abiity to colonize freshwater. In aquatic ecosystems, freshwater prey items generally lack polyunsaturated fatty acids, while marine prey items contain a lot. Therefore, increase in the copy number of this gene likely enables a more efficient use of freshwater food resources. Our data demonstrate that gene transposition of a key gene can increase ecological opportunities and trigger adaptive radiation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-253

Functional study of a polymorphic retrocopy in Drosophila

Jinbo Wang [1,*], Shengjun Tan [1], Yanan Mao [1], Yong Zhang [1]

[1]Key Laboratory of Zoological Systematics and Evolution & State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute Of Zoology, Chinese Academy Of Science, Beijng, China

Abstract: In fruit fly we previously assembled 15 polymorphic retrocopies and found that all retroposed loci are chimeras of internal retrocopies flanked by LTR retrotransposons. The internal retrocopies are often partially duplicated, and sometimes involving multiple parental genes. Interestingly, we found that LTR-mediated retrocopies are co-transcribed with their flanking LTR retrotransposons. Are any of these young retrocopies actually translated and functional? To answer this question we choose one gene X for in-depth studies. We first observed that X is highly transcribed in multiple tissues, including testis. By examining GFP knock-in flies which contains GFP reporter at the C-terminus of the coding sequence, we found that X is indeed translated in testis. Specifically, the fusion protein (X-GFP) is mainly detected in the spermatid nucleus during spermiogenesis process suggesting a regulatory role in male reproductive process. Therefore, with the generation of knock-out mutant, we are performing phenotyping, omics and genetic assays to illustrate the functionality of X in male reproductive process and how its functionality is achieved.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-249

The genetic basis of a key adaptive trait – radula genes in the radiation of Tylomelania

Leon Hilgers [1,*], Stefanie Hartmann [2], Michael Hofreiter [2], Thomas von Rintelen [1]

[1]Malacology, Museum für Naturkunde Berlin - Leibniz Institute for Evolution and Biodiversity Science, Berlin, [2]Institute of Biochemistry and Biology - Evolutionary and adaptive genomics, University of Potsdam, Potsdam, Germany

Abstract: Specializations of the feeding apparatus play a crucial role in numerous adaptive radiations. Examples include the beak of Darwin's finches as well as pharyngeal and oral jaws of East African cichlids. Despite intensive research, our understanding of the genetic evolution that underlies the diversification of such key adaptive traits is still in its infancy. *Tylomelania* is a genus of freshwater snails endemic to Sulawesi (Indonesia) that gave rise to adaptive radiations in the ancient lakes of the island. Lacustrine taxa developed unparalleled substrate-correlated radula (rasping tongue) diversity during trophic specialization which is believed to have driven these radiations. To investigate the genetic basis of this key adaptive trait, we generated morph-wise transcriptomes of mantle, foot and radula-forming tissue of *Tylomelania sarasinorum*, which exhibits a striking substrate-correlated radula polymorphism. Here we present the first radula-forming tissue transcriptome and compare sequence and expression information from both ecomorphs to investigate the genetic basis of radula shape and formation. Expression analyses illuminated tissue specific patterns and, combined with outlier scans, identified candidate genes that likely contribute to the evolution of radula diversity. This study is the first step towards uncovering the genetic basis of radula diversification in *Tylomelania* which will ultimately add to our understanding of the genetic underpinnings of adaptive traits.

Expanded summary*: Despite intensive research on the genetic basis of key traits in adaptation and speciation, which can be "key innovations", our understanding of the genetic underpinning of these traits is still in its infancy. The radula (rasping tongue) is an autapomorphy and key innovation of the Mollusca, central organ of the feeding apparatus and discussed to have evolved before the shell in early molluscs. Within gastropods the radula underwent remarkable adaptive evolution leading e.g. to the toxoglossan harpoon-like teeth of carnivorous turrids and conids.

Diversification of the feeding apparatus furthermore plays a pivotal role in numerous adaptive radiations. This is also true for *Tylomelania*, a genus of viviparous freshwater snails endemic to Sulawesi (Indonesia) which gave rise to well-known radiations in the ancient lakes of the island. Lacustrine clades of *Tylomelania* developed unparalleled substrate-correlated radula diversity. Here we present the first study to investigate the genetic basis of radula formation and identify candidate genes for radula shape divergence, which is an integral part of the trophic specialization believed to have driven the radiation of *Tylomelania*.

We focus on *Tylomelania sarasinorum*, a hard substrate dweller, which exhibits a pronounced habitat-correlated radula polymorphism and occurs on rock and wood substrates in Lake Towuti.

RNA-Seq was performed with four biological replicates of mantle, foot and radula forming tissue of both ecomorphs generating > 1 billion paired end reads. We conducted differential expression analyses between tissues and ecomorphs combined with allele frequency information and functional annotations to identify candidate genes for differences in radula shape. This study revealed a highly specialized expression profile of radula forming tissue. Gene expression differed substantially between the two biomineralizing tissues - mantle and radula, both of which have recently been in the focus of material scientists.

Identified candidate genes for radula shape will be tested with *in-situ* hybridization, aided by data on ultra-anatomy of radula forming tissue. Identification of genes contributing to the observed radula diversity is a crucial step in understanding key processes of this radiation at the genetic level, and will ultimately add to our knowledge about the genetic basis of adaptive traits from the population level up to key innovations on a phylum scale.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-247

Genome-wide analysis of the anticoagulant resistant Norway rat (Rattus norvegicus)

Sreyasi Biswas [1,*], Hans Michael Kohn [1]

[1]Biosciences, Rice University, Houston, United States

Abstract: In the post genome-sequencing era it is now feasible to re-examine classical examples of simple Mendelian adaptive traits, such as anticoagulant rodenticide resistance in the Norway rat (*Rattus norvegicus*). At enhanced resolution we discover that such traits in fact have a complex genetic architecture. In outbred natural populations, polygenic traits are probably evolving through modest changes in allele frequency at several loci, including standing variants, which were not easily mapped during earlier linkage analyses. In this study, we attempt to address the question of whether natural selection primarily acts on *de novo* mutations or on standing variations, or a combination of both, in anticoagulant resistant Norway rats from Northwestern Germany. Using whole genome sequencing data and restriction site digested marker sequencing (RADseq) collected from resistant and sensitive rats, we describe the population structure, the role of genetic drift, and the adaptive variants associated with resistance in that system. We compare the observed pattern to that of the ancestral population from China to examine the role of new mutations versus the role of ancestral standing variants. For the latter we include the genome sequence of *Rattus rattus*, the black rat. We use coalescent simulations and forward time simulations to test the odds of adaptive mutations to stand out, in population genetic terms, when compared to the genomic background that was affected by the recent demographic history of the Norway rat as it colonized the globe.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-246

Intraguild predation leads to a multitude of genomic changes in threespine stickleback

Sara Miller [1 2,*], Dolph Schluter [1], Marius Roesti [1]

[1]Zoology, University of British Columbia, Vancouver, Canada, [2]Neurobiology and Behavior, Cornell University, Ithaca, United States

Abstract: Biotic interactions among species are common and have been predicted to be an important mechanism for the evolution of phenotypic diversity. To date, the evolutionary effect of biotic selection upon an organism's genome remains largely unknown in natural populations. Following the last ice age, marine threespine sticklebacks (*Gasterosteus aculeatus*) became isolated in coastal freshwater lakes in British Columbia. Utilizing this unique natural history, we identified recently colonized lakes in which stickleback occur with and without prickly sculpin (*Cottus asper*), an intraguild predator of stickleback. Sculpin presence is linked to parallel changes in stickleback morphology and behavior. We report the results of an analysis examining the genome-wide effect of stickleback adaptation to sculpin. We generated a dataset of 5 million SNPs using whole genome re-sequencing of 23 populations. The main axis of genetic variation in freshwater populations is strongly associated with the presence of sculpin. To identify the regions of the genome that have differentiated in parallel and to quantify the strength of this divergence, we developed a principal component based genome scan method. Intraguild predation is associated with widespread but unevenly distributed selection across the genome. Despite the recent colonization of these lakes, evidence suggests the parallel occurrence of a large number of selective sweeps. Adaptation to intraguild predation may involve hundreds of genes with diverse functions. As a consequence, biotic interactions in the wild may lead to selection on a greater number of genes than previously appreciated.

Expanded summary*: Species interactions are ubiquitous in nature. Despite widespread interest in the evolutionary effect of biotic selection, how the genomes of wild populations evolve in response to selection from another organism is not well understood. A full comprehension of the genetic basis of adaptation to biotic selection requires measurements of the number, identity, genomic architecture, effect size, and source of genes under selection.

Quantifying the genome wide effect of biotic selection in the wild has been stymied by two major challenges. The first challenge is isolating an agent of selection in natural environments. Genomic comparisons between two environments are insufficient to measure the effect of biotic selection because putative signatures of selection can be confounded by local adaptation to other agents of selection (e.g. the abiotic environment). We attempt to address this challenge using the threespine stickleback (*Gasterosteus aculeatus).* The history of this species provides a natural experiment isolating the effect of a biotic agent of selection. Numerous freshwater lakes were formed in coastal British Columbia 10,000 years ago following the last ice age. Lakes were initially colonized by marine stickleback but quickly became isolated from each other as a result of isostatic rebound. An intraguild predator of stickleback, the prickly sculpin (*Cottus asper*), colonized a subset of these lakes. We identified study lakes in separate watersheds with similar abiotic traits, but which differed by the presence or absence of prickly sculpin. These populations provide replicated study of selection to a biotic agent, but at a longer time frame than is possible in experimental selection studies.

A second challenge is quantifying the effect of selection across the entire genome. Studies of wild populations have primarily utilized reduced representation genome scans (e.g. RADseq). These methods provide data for a limited portion of the genome and genetic differences not in linkage disequilibrium with markers will go undetected. We used whole genome re-sequencing of our study populations to quantify selection across most of the stickleback genome.

Prior studies have shown parallel phenotypic changes among stickleback populations that occur with sculpin including increased anti-predator defenses, a shift in diet, a narrower body shape, and a change in behavior. Matching these findings, we found strong, parallel genomic differentiation between lakes with and without sculpin. We developed a principal component based genome scan method to identify the regions of the genome that have differentiated in parallel between stickleback from lakes with and without sculpin. We observed parallel differentiation in nearly 2% of the genome. Outlier regions contained more than 500 genes with many diverse functions. Outlier regions were not randomly distributed with large portions of three chromosomes displaying elevated divergence between lakes with and without sculpin. Widespread differentiation among populations was further supported using hidden markov models to estimate the number of selective sweeps.

The presence of sculpin has had a rapid and profound effect on genomic divergence in stickleback. These findings suggest that biotic interactions may lead to selection on a greater number of genes than previously appreciated. If this amount of differentiation is typical for each agent of selections, a large proportion of organisms' genomes may be under selection.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-216

Evolution in the Constant Cold: Varied Loss of the Cellular Response to Heat Stress Among the Antarctic Notothenioid Fishes

Kevin Bilyk [1,*], Christina Cheng [1]

[1]Animal Biology, University of Illinois, Urbana Champaign, Urbana, United States

Abstract: Confined within the freezing Southern Ocean, the Antarctic notothenioid fishes have evolved to become highly cold-specialized. This specialization has come to impact many of their physiological and cellular responses to heat, including an apparent loss of the classic heat shock response (HSR). However, it remains unclear whether this HSR loss pervades the suborder and the extent to which such losses may pervade the broader transcriptional responses to heat stress. Understanding these losses is crucial in evaluating the adaptability of this stenothermal taxon that now faces the escalating threat of rising water temperatures due to Global Climate Change. In this study, we used RNA-Seq to assess the evolutionary status of these cellular responses to heat stress, comparing the transcriptional response to comparable, severe, heat stress in three select notothenioid lineages: the basal temperate *Eleginops maclovinus*, the nearest non-Antarctic sister species; the cryopelagic *Pagothenia borchgrevinki*, which inhabits Antarctica's iciest waters; and the highly derived *Chionodraco rastrospinosus*, one of the hemoglobin-lacking Antarctic icefishes. *E. maclovinus* continued to exhibit a robust heat shock response, supporting the HSR as a plesiomorphy that preceded the Antarctic notothenioid radiation. Examining the two Antarctic species, *C. rastrospinosus* maintained a robust response but lacked the classic HSR, while the *P. borchgrevinki* response was extraordinarily muted. These disparate transcriptional responses to heat stress suggest that the evolution of cold-specialization has progressed to differing extents among Antarctic notothenioid lineages, which may impact their capacity to adapt to a warming world.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-274

The genetic basis of variation in phenotypic plasticity

Karin R. L. van der Burg [1,*], James J. van der Burg [1], Robert D. Reed [1]

[1]Ecology and Evolutionary Biology, Cornell University, Ithaca, United States

Abstract: The ability to change phenotype in response to environmental conditions is prevalent throughout the tree of life, yet relatively little is known on how variation in phenotypic plasticity evolves. The Common Buckeye butterfly Junonia coenia often possesses strong seasonal wing color plasticity. We describe two J. coenia lines derived from plastic North Carolina individuals, selected for increased and reduced plasticity, respectively. Crossing individuals from these selection lines, we produced F3 individuals to map genomic loci associated with variation in a plastic response. Using whole genome resequencing data, we show that variation in a plastic response maps with high resolution to only five loci of interest containing a total of 4 Mb of sequence, which is less then 1% of the genome. We then identify the minimal genetic changes necessary for variation in plasticity using a combination of nucleotide variant association, gene expression differences, and chromatin accessibility variation between phenotypes. Within the five loci of interest, we investigate chromatin accessibility using ATACseq, and find that only a handful of accessible regulatory loci are variable between plastic and non-plastic individuals. We link these variable sites to differences in gene expression, thus identifying several candidate genes and regulatory loci for variation in plasticity. This work highlights the usefulness of combining novel genomic techniques with traditional mapping methods to identify narrow regions of interest associated with complex traits.

Expanded summary*: Phenotypic plasticity, or the ability to produce different phenotypes in response to environmental conditions, is one of the major drivers of phenotypic diversity, as it allows organisms to adapt to a wider range of different and/or changing environments. Despite widespread interest, relatively little is known on how variation in plasticity evolves. The butterfly *Junonia coenia* is readily selected for an increased or reduced plastic response, and it also shows natural variation in wing color plasticity in different populations, making it an ideal model to address both short and long term evolution of variation in plasticity. My work uses physiological (hormone signaling), molecular (gene expression, chromatin state) and genetic (mapping, population genetics) approaches to study (1) what developmental genetic mechanisms are involved in bringing about variation in plasticity, and (2) how do these mechanisms vary to allow reaction norm adaptation between natural populations?

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-266

Experiment on the long-term submerged cultivation of Podospora anserina: identification of acquired mutations

Ksenia Safina [1,*], Olga Kudryavtseva [2], Olga Vakhrusheva [1], Georgii Bazykin [1], Igor Mazheika [2], Ekaterina Budanova [2], Olga Kamzolkina [2], Alexey Kondrashov [3]

[1]Institute for Information Transmission Problems, [2]Faculty of Biology, Lomonosov Moscow State University, Moscow, Russian Federation, [3]University of Michigan, Ann Arbor, United States

Abstract: Ascomycete fungus *Podospora anserina* is frequently used as a model organism in studies on the mechanisms of ageing. While *P. anserina* has a limited lifespan with clear signs of ageing when grown on solid media, it shows no evidence of senescence when cultured under aeration on liquid media. Apart from transition to unlimited growth, submerged cultivation triggers pronounced changes in mycelium morphology. To determine whether there might be a genetic basis for the observed phenotypic changes we conducted a long-term submerged cultivation experiment of 8 replicate lineages of *P. anserina* and analyzed genetic changes that arose and became fixed in the replicate lines during the course of the experiment.
After 268 serial passages 8 replicate lines fixed a total of 129 point mutations, including 52 missense and 8 nonsense mutations. Numbers of both missense and nonsense variants were substantially higher than expected by chance.
The observed enrichment of the fixed mutations for protein-altering variants suggests that some of the mutations might be adaptive. Functional annotation of the fixed mutations provides further evidence that some of them might result in the adaptive phenotype: several of the fixed variants affect genes involved in growth and development (pro1, FlbA, FadA). Moreover, we observe parallel evolution at the gene level with 4 genes possessing protein-altering mutations fixed in parallel in more than one independent lineage.
This work was supported by the RFBR grant 16-04-01845a.

Statement: I am a sixth year student from the Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University. About three years ago, I joined the Laboratory of Evolutionary Genomics at the MSU and started to study problems related to the broad field of evolutionary biology. At the moment, I'm working on two unrelated projects: on the experimental evolution of *P. ancerina* and on the influenza A virus fitness prediction. And I would like to present at the SMBE meeting the first one. Being my first meeting outside my homeland, it would give me a great chance to discuss my work with more experienced scientists from around the world and to make new acquaintances in the scientific community.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-272

Antagonistic Pleiotropy under Genetic Drift

Mrudula Sunil Sane [1,*], Deepa Agashe [1]

[1]National Centre for Biological Sciences, TIFR, India, Bangalore, India

Abstract: Negative correlations between traits, or tradeoffs, may constrain trait evolution. Tradeoffs in microbial resource utilization shape ecological and biogeochemical processes. True tradeoffs stem from antagonistic pleiotropy (AP), when mutations are simultaneously beneficial for one trait and deleterious for another. Demonstrating such functional tradeoffs is difficult because strong selection on a trait allows unrelated genes to accumulate deleterious mutations, leading to erosion of those traits and causing spurious negative correlations across traits. To get accurate, unbiased estimates of the incidence of AP, we evolved replicate lines of *Escherichia coli* strains with varying ~2000-fold in mutation rates under minimal selection (mutation accumulation) for hundreds of generations. In this regime, populations evolved mainly under genetic drift. Evolved populations had 1–130 point mutations accumulated per line. We measured growth rate of each evolved line on different carbon sources to find instances of AP (increased growth on resource A and decreased growth on resource B). Across pairwise comparisons of resources, the proportion of replicate lines showing AP varied from 8% to 80%. Surprisingly, incidence of AP decreased with higher mutation rate and increased with greater metabolic dissimilarity between two resources. Unexpectedly, some of our strains repeatedly evolved novel resource utilization under genetic drift. Our study is the first to systematically quantify the incidence of functional AP and factors influencing it. Our work highlights the inherent variability and context dependence of AP, and suggests that evolution under genetic drift is a powerful method to reveal these patterns.

Expanded summary*: Antagonistic Pleiotropy under Genetic Drift

Mrudula Sane and Deepa Agashe

National Centre for Biological Sciences, TIFR, UAS-GKVK, Bellary Road, Bangalore 560065, India

The knowledge that organisms maximize resource allocation to one trait while reducing allocation to another has been around since before Darwin put forth his ideas on natural selection (Ferenci T, *Trends in Microbiology*, 2016). These negative correlations between traits are called tradeoffs. Tradeoffs govern diverse biological processes such as life-history strategies in insects (Sgro CM & Hoffmann AA, *Heredity*, 2004; Zera AJ & Harshman LG, *Annual review of Ecology and Systematics*, 2001) and ageing (Kirkwood TB, *Cell*, 2005). In bacteria, tradeoffs between resource-utilization traits determine composition of microbial communities, influence biogeochemical cycles and host-microbe interactions (Litchman E *et al*, *Frontiers in Microbiology*, 2015).

Tradeoffs are thought to constrain trait evolution by preventing the evolution of one "super-fit" generalist genotype, allowing the large diversity of microbial life on earth. Therefore, it is important to understand how tradeoffs arise.

Typically, tradeoffs are thought to be caused by antagonistic pleiotropy (AP), where mutations that are beneficial for one trait are directly deleterious for another. Although theoretical work predicts that AP is almost universal (Dillon MM *et al*, 2016), experimental evidence for AP comes entirely from lab populations of bacteria evolved under strong selection (Dillon MM *et al*, *Evolution*, 2016; Carroll SM *et al*, *Evolution*, 2014; Behrends V *et al*, *Molecular Biosystems*, 2014; Satterwhite RS *et al*, *Evolution*, 2015). But strong selection for one trait can lead to erosion of other traits by allowing deleterious mutations to accumulate, leading to spurious negative correlations between traits, and confounding our estimates of true tradeoffs.

To get unbiased and accurate estimates of AP, we evolved replicate populations of *E. coli* strains with varying mutation rates, under minimal selection (mutation accumulation) for hundreds of generations. Evolved lines had 1 to 130 accumulated point mutations per line. We measured growth rate of these lines on several different carbon resources, to identify incidence of AP (i.e increased fitness on resource A and decreased fitness on resource B). Across pairwise comparisons of resources, the proportion of replicate lines showing AP varied from 8% to 80%. Surprisingly, the incidence of AP decreased with higher mutation rate and increased with greater metabolic dissimilarity between the two resources. We also found that the proportion of synergistic decreases in fitness (i.e decreased fitness in both resource A and B) increased with mutation rate, and this trend was reversed for synergistic increases. Unexpectedly, we also found that some strains repeatedly evolved novel resource utilization under genetic drift. Our study is the first to systematically

quantify the incidence of functional AP and the factors that influence it. Our work highlights the inherent variability and context dependence of AP, and suggests that evolution under genetic drift is a powerful method to reveal these patterns.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP8

From candidate variants to fitness landscapes: investigating the rare origins of novel trophic specialists during adaptive radiation

Christopher Martin*

Abstract: A complete understanding of the adaptive phenotypes underlying species diversity requires investigation of both the ecological context in which they evolved and the diverse sources and histories of genetic variation contributing to these phenotypes. Here I integrate these perspectives in order to investigate the rare origins of adaptive radiation of three sympatric pupfish species endemic to a single Bahamian island, including a large-jawed scale-eating specialist and a molluscivore with a novel nasal protrusion. Jaw morphology in this clade is diversifying 1,000 times faster than neighboring islands, despite comparable levels of ecological opportunity and identical fish communities. This presents an outstanding mystery and opportunity to investigate the rare origins of novel specialist phenotypes in a system in which additional factors beyond ecology appear to constrain diversification. One possibility is the topography of the adaptive landscape: a complex fitness landscape with multiple peaks is driving the evolution of novel specialists on this island and generalist phenotypes are stranded on an isolated peak, which may explain the rare evolution of specialists. Alternatively, our population genomic and quantitative genetic studies indicate that only a few hundred SNPs out of 12 million underlie these species' phenotypes and only four moderate-effect loci underlie the major adaptive axis of jaw size, suggesting a tractable and highly localized genetic architecture. We are now tracing the histories of these candidate genomic regions and find evidence of both hard and soft selective sweeps within novel and well-characterized regulatory regions as well as a small role for adaptive introgression from distant islands in triggering adaptive radiation. Thus, it appears that the evolution of novel adaptive specialist phenotypes is triggered by either surprisingly subtle ecological shifts influencing the adaptive landscape or the fortuitous assembly of pools of standing genetic variation from diverse sources across the Caribbean.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-269

Environmental adaptation in House mice: the role of gene regulation along a latitudinal cline

Katya L. Mack [1,*], Mallory A. Ballinger [1], Megan Phifer-Rixey [2], Michael W. Nachman [1]

[1]Integrative Biology and Museum of Vertebrate Zoology, University of California, Berkeley, Berkeley, [2]Department of Biology, Monmouth University, West Long Branch, United States

Abstract: Geographic clines in allele frequencies are widely used as evidence for adaptation to spatially varying environments. Changes in *cis*- regulatory regions are thought to play a major role in the genetic basis of adaptation. However, few studies have linked cis- regulatory variation with environmental adaptation in natural populations. Here, using a combination of exome and RNA-seq data, we perform expression quantitative trait locus (eQTL) mapping and allele-specific expression analyses to study the genetic architecture of regulatory variation in wild house mice (*Mus musculus*) using 50 individuals from 5 populations collected along a latitudinal cline in eastern North America. We identify genes with clinally varying *cis*-eQTL where expression level is correlated with latitude. We also use weighted gene co-expression network analysis to identify sex-specific co-expression modules that are highly correlated with the clinally varying phenotype, body mass index (BMI). We identify candidate *cis*-eQTL associated with body mass variation in natural populations. These findings provide strong evidence for *cis*- regulatory elements as essential loci of adaptive clinal evolution in natural populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-225

The adaptive significance of natural genetic variation in the DNA damage response of Drosophila melanogaster

Nicolas Svetec*, Julie Cridland, Li Zhao, David Begun

Abstract: Despite decades of work, our understanding of the distribution of fitness effects of segregating genetic variants in natural populations remains largely incomplete. One form of selection that can maintain genetic variation is spatially varying selection, such as that leading to latitudinal clines. While the introduction of population genomic approaches to understanding spatially varying selection has generated much excitement, little successful effort has been devoted to moving beyond genome scans for selection to experimental analysis of the relevant biology and the development of experimentally motivated hypotheses regarding the agents of selection.
Solar UVB incidence is negatively correlated with latitude and an important agent of DNA damage in nature. Motivated by population genomic results, we investigated whether clinal variation in UVB incidence has led to genetic differences in the DNA damage response of *D. melanogaster* populations. Using a combination of population genomics, transcriptomics, and organismal phenotypic analysis, we show that genetic variation in the DNA damage response in *D. melanogaster* is maintained by spatially varying selection due to latitudinal variation in solar UVB incidence.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-224

Population Genomics of Parallel Adaptation to Cold Environments in Drosophila melanogaster

John Pool [1,*], Dylan Braun [1], Justin Lack [1]

[1]Laboratory of Genetics, University of Wisconsin, Madison, United States

Abstract: *Drosophila melanogaster* originated in tropical Africa before expanding into strikingly different temperate climates in Eurasia and beyond. Here, we show that elevated cold tolerance evolved not only as the species left Africa, but also twice within Africa, in the cool highlands of Ethiopia and South Africa. We sequenced more than 300 new inbred strain genomes from six natural population samples (encompassing a warm- and cold-adapted pair of populations for each origin of cold tolerance). Population genomic analysis, using our recently-described Population Branch Excess statistic (*PBE*) to assess evidence for local adaptation, then allowed us to assess evidence for genetically parallel evolution associated with recurrent cold tolerance adaptation. When *PBE* was applied to genomic windows (~4 kb), only limited evidence for parallel genetic differentiation of cold-tolerant populations was observed. In contrast, when we searched for single nucleotide polymorphisms (SNPs) with codirectional frequency change in two or three cold-adapted populations, strong genomic enrichments were observed from all comparisons. These findings could reflect an important role for selection on standing genetic variation leading to "soft sweeps". An intronic SNP at the synaptic gene *Prosap* showed a particularly strong pattern of parallel allele frequency change; and more generally, proteins involved in neurotransmission were enriched as potential targets of parallel adaptation. The ability to study cold tolerance evolution in a parallel framework will enhance this classic study system for climate adaptation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-220

Keeping it brief: evolution of a voltage-gated potassium channel in electric fish

Swapna Imani [1], Alfredo Ghezzi [2], Jason Gallant [3], Harold Zakon [1,*]

[1]Dept. of Neuroscience, The University of Texas, Austin, United States, [2]Dept. of Biology, University of Puerto Rico, San Juan, Puerto Rico, [3]Dept. of Zoology, Michigan State University, East Lansing, United States

Abstract: Nocturnally-active African Mormyrid electric fish emit and sense electric organ discharges (EODs) to detect objects around themselves and to communicate with conspecifics. Most mormyrids, which includes 200+ species, make an extremely brief EOD of a few hundred microseconds. A number of selection pressures have pushed EOD pulses to be unusually brief. Since voltage-gated potassium (Kv) channels determine action potential duration, we investigated whether Kv channels in the mormyrid EO evolved biophysical properties to shorten action potentials. We made transcriptomes from muscle and EO of a number of mormyrid species and identified a candidate Kv channel gene. The Kv channel gene, kcna7, expresses in vertebrate muscle and this gene duplicated in the teleost whole genome duplication. One paralog, kcna7b, retained muscle expression in mormyrids while the other, kcna7a, lost its expression in muscle and became expressed in the EO (which is a muscle derivative). Along with its change in expression, the mormyrid kcna7a gene underwent a burst of episodic diversifying selection whereas the kcna7a gene of non-mormyrid teleosts and the kcna7b gene of all sampled teleosts did not. We expressed kcna7a of a mormyrid (*Brienomyrus brachyistius*) and a non-mormyrid (*Gymnarchus niloticus*) in *Xenopus* oocytes and determined that the mormyrid kcna7a produces a current that activates more rapidly and at a more hyperpolarized membrane voltage, and with a steeper voltage-conductance curve than that of the non-mormyrid. Each of these three characteristics alone would shorten action potential duration; together, they would shorten it substantially. We noticed a patch of negatively charged amino acids in part (S3-S4 linker) of the mormyrid kcna7a channel that is unusual. We swapped this negative patch from the *Brienomyrus* channel with the neutral amino acids in the same position in the *Gymnarchus* channel and found that this was sufficient to reciprocally swap the biophysical properties of each channel onto the other. We suggest that the patch of negative amino acids changes the surface charge over the positively charged amino acids in the Kv channel voltage sensor (S4) allowing it to activate with less depolarization. We conclude that the teleost whole genome duplication event produced a duplicated Kv channel gene that underwent changes in expression and sequence resulting in its specialization for generating brief EOD pulses in a lineage of electric fish.
The authors gratefully acknowledge NSF grant IOS-1557857.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP13

Local adaptation of a human cold receptor along a latitudinal cline

Felix Key [1,2], Muslihudeen Abdul-Aziz [1,3], Roger Mundry [4], Benjamin Peter [5], Mauro D'Amato [6], Megan Dennis [7], Joshua Schmidt [1], Aida Andres [1,*]

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, [2]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany, [3]Australian Centre for Ancient DNA, The University of Adelaide, Adelaide, Australia, [4]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [5]Department of Human Genetics, University of Chicago, Chicago, United States, [6]BioDonostia Health Research Institute, San Sebastian, Spain, [7]Genome Center, University of California, Davis, Davis, United States

Abstract: Ambient temperature is a critical environmental factor, and persistent cold likely a major challenge as early modern humans first colonized high-latitude non-African environments. A few local adaptations have been identified in subarctic human populations, mostly involving metabolic changes in response to local diets. As of yet, though, we know of no genetic adaptation to temperature perception. Using a battery of tools we show that *TRPM8*, the gene that encodes the TRP cation channel that mediates cold perception, evolved under local positive selection in populations living at moderate to high latitudes. The derived allele of upstream SNP rs10166942 shows unusually high population differentiation, with frequencies that range from 5% in Nigeria to 88% in Finland; in fact, its frequency correlates significantly with latitude and temperature beyond what can be explained by population history, both in the 1000 Genomes and the SGDP datasets. A sophisticated Bayesian approach revealed that the allele was neutral in Africa and, after the out-of-Africa bottleneck, it was differentially targeted by positive selection across Eurasian populations. Interestingly, rs10166942 ancestral allele has been associated with protective effects from migraine, a debilitating disorder that has its highest prevalence in individuals of European descent. Local adaptation on previously neutral standing variation may thus have influenced not only temperature perception, but also the risk of pain phenotypes in certain human populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-221

Plasticity first: molecular signatures in a complex morphological trait

Tal Dagan*

Abstract: Filamentous cyanobacteria that differentiate multiple cell types are considered the peak of prokaryotic complexity and their evolution has been studied in the context of multicellularity origins. Cyanobacteria strains that form true-branching filaments exemplify the most complex cyanobacteria. The mechanisms underlying the true-branching morphology remain poorly understood, despite of several investigations that focused on the identification of novel genes or pathways. An alternative, highly debated, route for the evolution of novel traits is based on existing phenotypic plasticity. According to that scenario, the fixation of a novel phenotype precedes the fixation of the genotype. Here we show that the evolution of transcriptional regulatory elements constitutes a major mechanism for the evolution of new traits. We identified conditions that reconstitute the ancestral branchless phenotype of two true-branching *Fischerella* species and compared the transcription start sites (TSSs) between the two phenotypic states. Our analysis uncovers several orthologous TSSs whose transcription level is correlated with the true-branching phenotype. These TSSs are found in genes that encode for septosome and cell division components (e.g., *fraC* and *mreB*). Our results reveal a route for the evolution of the true-branching phenotype in cyanobacteria via modification of the transcription level of pre-existing genes. Our study supplies evidence for the plasticity-first hypothesis and highlights the importance of transcriptional regulation in the evolution of novel traits. The approach we used presents the opportunity to investigate the role of phenotypic plasticity in evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-223

Phenotypic signatures of adaptation in maize landraces provides a detailed look selection operates across entire landscapes

Daniel Gates [1,*], Jeffrey Ross-Ibarra [1]

[1]UC Davis, Davis, United States

Abstract: The process of adaptation has been a focal point of evolutionary biological studies for centuries. It has been well established that adaptive patterns are often confounded by unmodeled demography, ancestry, or gene flow. What is less well understood, however, is the dependence between adaptation and the genetic architecture of the phenotypes under selection and the spatial scale of the selective pressures. Open pollinated landrace maize plants in Mexico and South America are an excellent system to examine the genetic basis of adaptation across landscapes because they have undergone rapid changes that are consistent with adaption to the challenges of growing across a wide array of environments. Recent genetic work suggests that elevational adaptation may have targeted traits with both simple and complicated genetic architecture in different geographic regions. This provides an opportunity to examine how variable adaptation is across entire landscapes and genomes. We present analyses from over 2,000 landrace accessions of maize genotyped at over 600,000 SNPs and confirm a strong overall genomic signature of adaptive responses to elevation and that adaptation is driven in large part by major effect inversions. By using newly developed spatially explicit methods we are also able to identify the spatial scale that adaptation occurs at across the landscape. We supplement our results of outlier locus analyses with genetic and phenotypic data from previous GWAS panels to show that while large effect changes like inversions contribute to adaptive phenotypes, there is still a substantial amount of adaptive phenotypic change that driven by genome wide changes in small effect loci which are often missed by traditional methods. Our research shows how genomic data implemented in large landscape level sampling can give unprecedented resolution in our ability to detect patterns of adaptation across entire landscapes. Additionally, our results underscore the importance of developing phenotypic explicit methods that provide compliments to the traditional genetic signatures of adaptation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

OTH-MP9

Natural selection drives rapid neofunctionalization of Drosophila duplicate genes

Xueyuan Jiang [1,*], Raquel Assis [1]

[1]Biology, Pennsylvania State University, University Park, United States

Abstract: Gene duplication is thought to play a major role in the evolution of phenotypic novelty. Though recent studies have shown that duplicate genes experience rapid protein sequence and expression divergence, little is known about the underlying mechanisms. Here, we use population-genetic inference to elucidate the types, targets, and timing of evolutionary forces acting on recently duplicated genes in *Drosophila melanogaster*. Consistent with previous studies, we find evidence of elevated protein sequence evolutionary rates in young (child), but not in old (parent), duplicate gene copies. Further analyses indicate that accelerated evolutionary rates can often be attributed to natural selection, rather than to relaxation of selective constraint. We also show that types of natural selection differ between duplicates with conserved and diverged gene expression profiles, as well as between protein-coding and noncoding regions of these genes. Moreover, we demonstrate that selection is often strongest immediately after duplication and weakens over time, and that duplicates that underwent recent strong selective sweeps tend to have spermatogenesis-associated functions. Together, our findings reveal that natural selection acts strongly, swiftly, and specifically on young *Drosophila* duplicate genes, driving their rapid functional evolution and perhaps contributing to the origin of phenotypic innovation.

Expanded summary*: Gene duplication creates two copies of an ancestral gene—an orthologous "parent" copy and a new "child" copy. According to evolutionary theory, functional redundancy of duplicate genes results in relaxation of selective constraint in one copy, nearly always leading to an accumulation of deleterious mutations and pseudogenization of that copy within a few million years (Ohno 1970; Lynch and Conery 2000). Yet functional duplicate genes are abundant in the genomes of organisms from all three domains of life (Zhang 2003), can be hundreds of millions of years old (Long *et al.* 2003), and often quickly become essential for viability in *Drosophila* (Chen *et al.* 2010). These observations suggest that many duplicate genes manage to escape pseudogenization and be retained over long evolutionary time periods.

The sheer abundance, long-term retention, and functional importance of many duplicate genes has led to widespread interest in their evolution and how they contribute to phenotypic innovation. Evolutionary studies have shown that duplicate genes experience rapid sequence and expression divergence (e.g., Gu *et al.* 2002; Wagner 2002; Makova and Li 2003) that is often asymmetric and increased preferentially in child copies (e.g., Wanger 2002; Gu *et al.* 2005; Assis and Bachtrog 2013). Moreover, a recent expression analysis of *Drosophila* duplicate genes found evidence for neofunctionalization, or acquisition of novel functions, in a majority of child copies (Assis and Bachtrog 2013). However, despite these exciting findings, little is known about the evolutionary forces driving the sequence and functional evolution of duplicate genes.

In our analysis, we use population-genetic inference to interrogate the evolutionary forces underlying the functional divergence of duplicate genes that arose in *D. melanogaster* after it split from *D. pseudoobscura*. Specifically, we are interested in understanding the role of natural selection in this process, as well as its targets, strength, and timing after duplication. To address these problems, we utilize sequence data from 12 *Drosophila* species (FlyBase Consortium 2017) and polymorphism data from *D. melanogaster* (Mackay *et al.* 2012) to conduct several sequence-, tree-, and haplotype-based tests for selection. Thus, our study utilizes recently available large-scale genomic data to answer an age-old question in evolutionary genomics: how do novel functions arise after gene duplication?

Consistent with previous studies, we find that protein evolutionary rates are elevated in child, but not in parent copies. However, examination of polymorphism and substitution data reveals that increased rates are often due to positive selection, rather than to relaxed constraint. These analyses also show that selective forces differ between functionally conserved and neofunctionalized child copies, as well as among the exons, introns, and UTRs of these genes. Moreover, a tree-based method indicates that selection is often strong immediately after duplication and becomes weak or absent over time. Last, a haplotype-based test identifies many child genes

that underwent recent selective sweeps, and gene ontology analysis uncovers an overrepresentation of spermatogenesis-related functions in these genes.

Hence, our research demonstrates that rapid evolution of young *Drosophila* duplicate genes is often driven by natural selection, rather than by relaxed selective constraint. We also show that selective pressures differ between functionally conserved and neofunctionalized genes, as well between their protein-coding and noncoding regions. Moreover, our findings suggest that selection is most intense immediately after duplication and weakens over time. Finally, we reveal that young duplicates that underwent recent strong positive selection are often involved in spermatogenesis. Together, our results elucidate the importance and mechanisms of natural selection in young *Drosophila* duplicate genes, providing a general framework for understanding how novel functions arise by gene duplication.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-261

Molecular and evolutionary origins of seasonal camouflage in snowshoe hares

Matthew Jones [1,*], José Melo-Ferreira [2], Paulo Célio Alves [2], Colin Callahan [3], Diana Lafferty [4], Jeffrey Jensen [5], L. Scott Mills [6], Jeffrey Good [1]

[1]Department of Organismal Biology, Ecology, & Evolution, University of Montana, Missoula, United States, [2]CIBIO, Universidade do Porto, Vairao, Portugal, [3]University of Montana, Missoula, [4]College of Natural Resources, North Carolina State University, Raleigh, United States, [5]School of Life Sciences, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, [6]College of Forestry and Conservation, University of Montana, Missoula, United States

Abstract: Linking genotypes to adaptive phenotypes in nature is a fundamental goal in evolutionary biology. Several species of mammals undergo seasonal color molts as part of a coordinated suite of responses to seasonally varying environments. Yet the genetic changes that cause dynamic seasonal shifts in pelage coloration, and seasonal plasticity in general, remain largely unknown. In snowshoe hares (*Lepus americanus*), autumn molts to white winter coats are cued by photoperiod and closely track the local onset of snow cover. However, hares in mild coastal climates with ephemeral snow cover have secondarily evolved brown winter coats. We have combined population genomics, association mapping, pedigree analyses, and functional genomic studies to resolve the molecular and evolutionary basis of alternative winter white or brown pelage morphs. Field estimates of coat color-related survival combined with genomic signatures of positive selection at causative coat color alleles reveal strong natural selection on the maintenance of locally adaptive seasonal crypsis. Additionally, we identified a strong candidate causal mutation that suggests an intriguing role of regulatory network co-option in the evolution of seasonal color molts. Our results thus provide one of the first examples of a clear genotype-to-phenotype-to-fitness link for a seasonally changing trait and insights into the molecular mechanisms that may be important in the evolution of complex adaptations to seasonal environments.

Expanded summary*: A complete understanding of adaptation requires resolving both the ultimate and proximate causes of adaptive trait evolution. Towards this end, we must identify and measure the selective forces acting on phenotypic variation and dissect the specific genetic changes and functional mechanisms underlying this variation. By linking genotype, phenotype, and fitness we can address fundamental, long-standing questions about adaptation. Is adaptation often convergent, using the same genes or even mutations across diverse taxa? What is the origin of adaptive variation? What molecular mechanisms underlie the evolution of phenotypic complexity? Though considerable progress has been made in identifying genetic targets of positive direction selection, the task of linking genes and mutations to ecologically relevant traits is difficult. Establishing the causal relationships among genotype, phenotype, and fitness has therefore remained one of the greatest challenges in evolutionary biology.

Animal coloration provides a rich set of traits for linking genotype, phenotype, and fitness because the molecular pathways and selective mechanisms underlying pigmentation are generally well known. Indeed, studies of constitutive forms of animal pigmentation have contributed fundamental advances in our understanding of the genetic basis of adaptive evolution. One surprising insight gained from studies of adaptive animal coloration was the repeated role of the same genes across diverse taxa, so called 'genomic hotspots' of adaptation. In mammals, birds, and reptiles, major genomic hotspots for adaptive coloration are *Mc1r* and *Agouti*. While identification of these 'genomic hotspots' has offered surprising insight into the repeatability of the evolutionary process, it remains unclear whether some of the principles revealed from studies of simple coloration phenotypes can be extended to more complex scenarios, including the evolution of phenotypically plastic coloration traits.

Many organisms have evolved mechanisms of plasticity for numerous behavioral, physiological, and morphological traits that allow them to cope with predictable seasonal environments. As part of this integrated suite of responses to seasonally changing environments, many birds and mammals evolved circannual color molts likely to track local changes in snow cover. While the ecology and physiology of seasonal coat color change has been the subject of over a century of intense research, the genetic changes that cause seasonal shifts in pelage coloration, and seasonal phenotypic plasticity in general, have remained largely unresolved. In

wild snowshoe hare populations, a striking link exists between coat color change and fitness as mismatched hares suffer substantial increases in predation. However, hares in mild coastal climates with ephemeral snow cover have secondarily evolved brown winter coats. Our research has focused on understanding the genetic basis of geographic variation for white versus brown winter coats in snowshoe hares (*Lepus americanus*). Using a genome-wide association approach, we have shown that this classic color polymorphism is caused by genetic variation at the 'hotspot' pigmentation gene *Agouti*. Population genomic analyses have further revealed that the evolution of brown winter coats in mild coastal climates involved positive directional selection on an introgressed *Agouti* allele from the non-seasonally changing black-tailed jackrabbit (*Lepus californicus*). We discovered a ~1 kb insertion-deletion variant in a cis-regulatory region of *Agouti* that perfectly associates with winter coat color. The deletion is associated with brown winter coats and appears to represent a reversion to the ancestral state found in rabbits and other non-changing mammals. These patterns suggest that the gain of this candidate cis-regulatory element (CRE) may have also played a crucial role in the origin of seasonal coat color change in snowshoe hares. The candidate *Agouti* CRE is also found in cis-regulatory regions of at least two other genes involved in hair growth cycles, indicating that it may play a general role in the coordinated regulation of hair growth and pigmentation development during seasonal color molts.

These results represent the first gene-level dissection of a seasonally changing trait that has been directly linked to fitness. We show that complex seasonal coloration traits evolved through changes at the same genes as simple, static coloration traits. Furthermore, our research provides one of the only examples of adaptive trait introgression in mammals. Finally, our research suggest that the co-option of existing regulatory networks may be an important mechanism underlying the evolution of novel phenotypic traits, particularly as a mechanism to coordinate highly integrated biological processes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-259

Genetic divergence and phenotypic plasticity in yeast strains subjected to repeated fly-gut vectoring.

Kelly M. Thomasson [1,*], Stephen R. Proulx [1]

[1]Ecology, Evolution and Marine Biology, University of California, Santa Barbara, United States

Abstract: Yeasts are a potentially useful group of organisms for studying of community dynamics, divergence and speciation: they are subjected to a large range of selective pressures, exhibit evidence of rapid speciation, yet also coexist in sympatry with closely related sister species. There is substantial evidence that sister species–within the genus *Saccharomyces*–have diverged in sympatry, but little is known about the mechanism by which these species diverged. Previous research has shown that spores have higher survival rates than vegetative yeast cells when both are passaged through insects. Is selection to survive insect guts sufficient to cause divergence and speciation in the genus *Saccharomyces*? We performed a long-term evolution experiment in which we subjected strains of the yeast, *S. cerevisiae*, from five different global regions, to the repeated selective pressure of ingestion and digestion by the fruit fly, *Drosophila melanogaster*. We found that evolved yeast strains have increased their rates of sporulation relative to those of ancestral strains, and these changes were quantified by the strain's ability to vector through insects. To substantiate the degree to which these phenotypic changes are due to genetic divergence or phenotypic plasticity, we assessed the competitive abilities and the survivorship through the insect gut of the evolved and control strains after one and five days of relaxed fly-gut selection pressure. Ongoing work to sequence the evolved strains will aim to determine the precise genetic basis of this adaptation within a model organism capable of rapid and prolific divergence.

Expanded summary*: The mechanisms of sympatric speciation are still poorly understood. Particularly when the reproductively isolating mechanism is not known to be instantaneous, it is unclear how incipient species genetically separate and become isolated while not competitively excluding each other. An Ecological trade-off such as the sporulation rate trade-off found in yeasts is a mechanism that not only promotes coexistence of two interacting incipient species, but may also promote divergence and speciation. Evidence of divergence is especially evident in yeast communities that are frequently subjected to ingestion and digestion by insects such as the fruit fly, *Drosophila melanogaster*. Here the ability to rapidly reproduce acts as a trade-off with the ability to survive passage through the fly gut. Within one yeast community there is substantial variation between individual sporulation rates and it is thus, sporulation rate within a yeast community is subject to selective pressures. To what degree is this selective pressure due to insect ingestion and digestion? Is the selective pressure of insect digestion sufficient to cause divergence and speciation in the yeast, *Saccharomyces cerevisiae*?

Colonization of a new habitat by vegetative cells and spores follows a markedly different trajectory such that the phenotypic life-history state (vegetative or spore) determines characteristics such as gene flow, growth rate and survival. Both competitive ability and genetic divergence are important factors in these microbial communities but plasticity is also known to play a substantial role in expressed phenotypes in communities experiencing rapid and unpredictable environmental changes.

Results from this study offer evidence of the degree to which these phenotypic changes are due to genetic divergence, epigenetic factors or phenotypic plasticity. Plasticity may play a role in divergence, but reproductively isolating mechanisms must be genetically fixed. For this reason, plasticity, epigenetic modifications and genetic mutations are often considered to be on a spectrum of permanency. Evidence of plasticity may indicate that these incipient species are still diverging or that phenotypic divergence is not indicative of speciation. Previous work has described reproductively isolating mechanism within metabolic pathways. Sequencing of key loci known to contribute to sporulation or with putative physiological functions may also inform the scientific community of more-likely isolating mechanisms within a model organism capable of rapid and prolific divergence.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-257

A little more DNA, a little less pigment: a CNV associated with reduced pigmentation in domestic pigeon

Rebecca Bruders [1,*], Michael Shapiro [1 2], Edward J. Osborne [2], Zev Kronenberg [2], Mark Yandell [2]

[1]Biology, [2]Human Genetics , University of Utah, Salt Lake City, United States

Abstract: The domestic rock pigeon (*Columba livia*) is a compelling model to understand molecular mechanisms of diversity because this species consists of over 300 different breeds with spectacularly variable phenotypes. One derived trait, "almond," is characterized by random sprinkling of pigmented and apigmented regions within and among feathers throughout the body. Classical genetic studies suggest almond is caused by a dominant sex-linked mutation. Additionally, homozygous almond males (ZZ sex chromosomes) develop severe eye defects and lack pigmentation, whereas hemizygous almond females (ZW), which lack a wild-type copy of the almond allele, do not develop these defects. This suggests that dosage of the mutant allele, rather than absence of the wild-type allele, is responsible for these eye phenotypes. We compared the genomes of almond pigeons to non-almond pigeons to identify a candidate almond locus on the Z chromosome. We found a substantial increase in sequencing coverage in this region in almond birds, indicative of a copy number variant. There are 5 genes in this region including *MLANA*, a melanosome maturation gene. No fixed coding changes were found in the genes in this region, indicating that this copy number increase is associated with the sprinkling of plumage pigmentation and eye defects in almond pigeons. By identifying the mechanisms responsible for almond-associated phenotypes, this work will provide key insights into pigment variation, eye development, and the connection between these two processes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-264

Assortative mating and the dynamical decoupling of genetic admixture levels from phenotypes that differ between source populations

Jaehee Kim [1,*], Amy Goldberg [1], Michael Edge [1], Noah Rosenberg [1]

[1]Biology, Stanford University, Stanford, United States

Abstract: Admixed populations descend from two or more source groups that have long been separated, and that therefore might have possessed distinct patterns of genotype and phenotype at the beginning of the admixture process. Frequent examples exist, however, in which genotypes or phenotypes initially associated with ancestry in one source population are decoupled from overall admixture levels, so that they no longer serve as proxies for genetic ancestry. We develop a mechanistic model for describing the joint dynamics of admixture levels and phenotype distributions in an admixed population. The approach includes a quantitative-genetic model that relates a phenotype to underlying loci that affect its trait value. We consider three forms of mating. First, individuals might assort in a manner that is independent of the overall genetic admixture level. Second, individuals might assort by a phenotype that is initially correlated with the genetic admixture level. Third, individuals might assort by the genetic admixture level itself. Under the model, we explore the relationship between genetic admixture level and phenotype over time, studying the effect of the genetic architecture of the phenotype. We find that the decoupling of genetic ancestry and phenotype can occur surprisingly quickly, especially if the quantitative phenotype is driven by a small number of loci. We also find that assortative mating attenuates the process in relation to a scenario in which mating is random with respect to genetic admixture and with respect to phenotype.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-262

Allelic diversification at the self-incompatibility locus in Arabidopsis

Maxime Chantreau [1], Thomas Brom [1], Celine Poux [1], Xavier Vekemans [2], Sylvain Billiard [1], Vincent Castric [1,*]

[1]CNRS, UMR 8198 -Evo-Eco-Paleo, Univ. Lille, [2]CNRS, UMR 8198 -Evo-Eco-Paleo, Univ. Lille,, F-59000 Lille, France

Abstract: Self-incompatibility in plants of the Brassicaceae family is controlled by a highly diversified molecular lock-and-key system consisting of a large set of specific haplotypic combinations of two tightly linked genes : *SCR*, which is expressed in the anther tapetum and encodes the male recognition specificity and SRK, which is expressed at the stigma surface and encodes the female recognition specificity. This system has been a textbook example of natural (balancing) selection, in the form of a strong reproductive advantage for individuals expressing rare alleles. While the many highly divergent haplotypes segregating is one of the most defining features of self-incompatibility genes, the question of how so many lock-and-key combinations could arise raises a series of interesting theoretical and mechanistic problems. In particular, the emergence and evolutionary success of any novel haplotype from an ancestral form entails at least two individual mutations, one on each of the two genes. In this talk, I will first detail how we are investigating the conditions under which the fitness valley represented by single mutants can still be crossed in spite of the fitness penalty entailed by lack of self-recognition and ensuing inbreeding depression. I will then detail a set of ongoing experiments based on ancestral resurrection of ancient self-incompatibility genes in the plant Arabidopsis thaliana to put these theoretical predictions directly to the test and try and catch the diversification process in flagrante delicto. Overall, our data using this simple and experimentally tractable biological system provide insight into the broader issue of how functional and regulatory novelty can arise in natural populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-254

FLOWERING LOCUS T2 is mediating local adaptation in a key life history trait in European aspen

Jing Wang [1,*], Pär Ingvarsson [2]

[1]Centre for Integrative Genetics (CIGENE), Norwegian University of Life Sciences, Ås, Norway, [2]Department of Plant Biology, Swedish University of Agricultural Sciences , Uppsala, Sweden

Abstract: Local adaptation plays a fundamental role in many plant species, but the genetic architecture of locally adapted traits remains poorly understood. Here we integrate multiple sources of data to identify a ~700kb region on chromosome 10 that mediate local adaptation in a key life-history trait, growth cessation, in *Populus tremula* across a latitudinal gradient in Scandinavia. This region harbours a *P. tremula* homolog of *FLOWERING LOCUS T*2 (*PtFT*2) that has long been implicated in the regulation of growth cessation and dormancy induction in perennial plants. We identify a recent selective sweep that is centred on the *PtFT*2 gene and that is restricted to the northernmost populations, suggesting that high-latitude populations have undergone strong and recent adaptation in response to post-glacial colonization of northern environments. Our results provide empirical evidence that the genomic architecture of local adaptation in high-gene flow species enrich for a few large-effect and/or tightly clustered loci rather than many independent loci of small effect.

Expanded summary*: The initiation of growth cessation and dormancy late in the growing season represents a critical ecological and evolutionary trade-off between survival and growth in most perennial plants. Growth cessation and dormancy induction are critical processes for the development of cold hardiness but developmental processes resulting in complete endodormancy require weeks to complete and this reduce the length of the season during which active growth can take place. Current evidence suggest that most important environmental cues regulating the initiation of dormancy in perennial plants growing at northern latitudes are a shortening of the photoperiod although exposure to extended periods of low, nonfreezing temperatures may also contribute. Adaptation to large-scale environmental gradients that shape the length of the growth are therefore crucial in perennial plants such as long-lived forest trees as individuals persist in environments long periods of time and cannot easily migrate to more benign physical conditions in the face of a changing climate. While local adaptation is well documented at the phenotypic level in many perennial, the genetic architecture of locally adaptive traits remains poorly understood.

In this study, we analysed whole genome re-sequencing data from 94 unrelated *P. tremula* (European aspen) trees sampled from twelve populations spanning a latitudinal gradient of c. 20 latitude degrees across Sweden. By taking advantage of both population genetics and association-mapping approaches, our results provide convincing evidence that a single c. 700 kbp region including the candidate gene *PtFT2* constitutes a locus with a major effect on a locally adapted trait, the timing of bud set, in Swedish populations of *P. tremula*. Furthermore, we find that changes in gene expression in responsible to day length is consistent with *PtFT2* mediating local adaptation, and RNAi mediated down-regulation of *PtFT2* mimics differences in grow cessation under field conditions. All these results further suggest that the *PtFT2* locus is likely to be the candidate gene in regulating the local adaptation of European aspen in Sweden.

In particular, we identify a recent and strong selective sweep that is regionally restricted to the northernmost populations. It indicates that high-latitude populations likely have undergone a stronger adaptive response to the greater environmental perturbation during the post-glacial colonization of northern Scandinavia.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-270

A key to coral success: Molecular Evolution of the genus Acropora

Groves Dixon*, Mikhail Matz

Abstract: The genus *Acropora* is the richest and most morphologically diverse of all Scleractinian genera. This exceptional success is thought to depend on a key innovation. Unlike other Scleractinia, Acropora corals have two polyp types, axial and radial polyps, that facilitate a division of labor within the colony. Axial polyps are located at branch tips and are primarily responsible for growth, while radial polyps along the sides of branches are responsible for heterotrophic feeding and reproduction. What were the molecular evolutionary events that lead to unique and highly adaptive phenotype? To address this question, we assembled a set of over 5,000 orthologous sequences from 23 Anthozoan transcriptomes and used PAML to test for evidence of positive selection in the lineage preceding Acropora diversification. To identify genes likely to be involved in development of dimorphic polyps, we cross-reference genes showing evidence for positive selection with a gene expression dataset comparing branch sides and branch tips in Acropora corals.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of phenotype evolution

POB-267

Gene expression level insights of asexuality in Daphnia pulex

Zhiqiang Ye*, Xiaoqian Jiang, Michael Lynch

Abstract: Asexual lineages have been suffered from selective disadvantages due to the lack of beneficial mutations, yet they are still geographically widely distributed. The genetic basis and underlying mechanisms of asexuality are mostly unknown. Here in this study, we address this issue by contrasting the genome-wide gene expression profile between cyclical parthenogenesis (CP) and obligate parthenogenesis (OP) *Daphnia pulex* lineages. 915 genes were identified to contain SNPs that present in all OP lineages, but none of the SNPs was found in CP lineages. 167 of these genes showed divergent expression between asexual and sexual lineages. Moreover, we found 257 out of 915 genes showed allele-specific expression between the two alleles. Also, we showed unique transposable element insertions in sexual/asexual lineages that can alter the expression of adjacent genes, which provides strong candidates regions/genes for future asexuality analysis.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-70

Adaptive Evolution of Avian RIG-I-Like Receptors

Wanjing Zheng [1,*], Yoko Satta [1]

[1]Evolutionary Studies of Biosystems, SOKENDAI, Hayama, Japan

Abstract: Innate immunity is an evolutionary ancient defense mechanism and is highly conserved across eukaryotes, in which a variety of PRRs (pattern recognition receptors) recognize pathogens and trigger signaling pathways modulating inflammatory reaction and adaptive immunity. Vertebrates have acquired the adaptive immunity, while innate immunity remains the defense system at the first line against pathogens. Birds resemble mammals in the respect of possessing both innate and adaptive immunity but carry much less proportion of EVEs (endogenous viral elements) in their genomes. We hypothesize that the difference in viral insertion frequency is related with the difference in the function of some anti-viral PRRs. RIG-I like receptors (RIG-I, MDA5, and LGP2) are intracellular PRRs for non-self RNA. We performed evolutionary analysis on RIG-I like receptor orthologs in birds. Results indicate that avian RIG-I shows higher pseudogenization frequency than MDA5 and LGP2, and that MDA5 has evolved under less diverged selection or higher functional constraint than the mammalian ortholog.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-71

Structural Reconstruction of Carbon Metabolism Before LUCA

Isabela Jeronimo Bezerra Marcos [1,*], Savio Torres de Farias and Laboratorio de Genética Evolutiva Paulo Leminski

[1]Universidade Federal da Paraiba, Joao Pessoa, Brazil

Abstract: Amongst the modern metabolic pathways, parts of the Carbon Pathway is believed to have been present in or even before the Last Universal Common Ancestor. Storey et al (2004) suggests that the first enzymes worked as binding domains, simply increasing the probability of the reaction. Farias et al (2016) reconstructed the ancestral sequences of tRNAs and found similarities with enzymes that could have been present in LUCA. Enzymes from the glycolytic and pentose phosphate pathways were present on their results. Here we propose the structural reconstruction of the metabolic network present at this very early stage of life, based on the proposed proteome. We also suggest the early steps of their evolution based on phylogenetic analysis of the studied proteins and show which and how substrates probably reacted to the polypeptides. We align the obtained ancestral structures with homolog modern proteins and compare the structural phylogeny with phylogenies based in sequences to understand the evolutionary history of this metabolic pathway.

All of the work was made in silico, now we are going to test the docking in laboratory by chemically producing the calculated molecules. We understand that our results confirm the idea that the first enzymes would have decreased the entropy of the system, and that tRNAs could have worked as the first genes. We posit that the development of the early translating machinery created the possibility of the evolution of the first proteins. Systems which had proteins that increased the efficiency of pre-existing pathways were positively selected.

Statement: I am on my last semester of undergraduate, for which I will receive a diploma in Biological Sciences from the Universidade Federal da Paraíba, in the state of Paraiba - Brazil. I have been committed to deepening my knowledge in Evolutionary Biology since high school and I have been engaged with research activities since my first semester at university. So far I have developed three research projects, and I have been part of two others, all on topics related to evolution but answering questions in distinct fields such as Human Developmental Genetics, Biochemistry, Molecular Evolution, Morphometrics, etc.

After having the previously mentioned opportunities, I can say with certainty that I expect to continue researching. I am especially fascinated by molecular evolution. However, there are many fields within Molecular Biology and Evolution which I am still not familiar with. Thus, I understand that going to the Annual Meeting of the Society for Molecular Biology and Evolution would be extremely enriching for first of all expanding my knowledge on these topics, and maybe find new possibilities for graduate school.

Secondly I see conferences as an opportunity to exchange knowledge with other young and future scientists. I am hopeful I will receive useful feedback for my presentation, which will enhance my work, and that I will be able to contribute with ideas for others. Finally, I am obviously excited to attend the symposia and meet the speakers.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-76

Secreted proteins defy the expression level–evolutionary rate anticorrelation

Felix Feyertag [1], Patricia Berninsone [1], David Alvarez-Ponce [1,*]

[1]Department of Biology, University of Nevada, Reno, Reno, United States

Abstract: The rates of evolution of the proteins of any organism vary across orders of magnitude. A primary factor influencing rates of protein evolution is expression. A strong negative correlation between expression levels and evolutionary rates (the so-called E−R anticorrelation) has been observed in virtually all studied organisms. This effect is currently attributed to the abundance-dependent fitness costs of misfolding and unspecific protein–protein interactions, among other factors. Secreted proteins are folded in the endoplasmic reticulum, a compartment where chaperones, folding catalysts, and stringent quality control mechanisms promote their correct folding and may reduce the fitness costs of misfolding. In addition, confinement of secreted proteins to the extracellular space may reduce misinteractions and their deleterious effects. We hypothesize that each of these factors (the secretory pathway quality control and extracellular location) may reduce the strength of the E–R anticorrelation. Indeed, here we show that among human proteins that are secreted to the extracellular space, rates of evolution do not correlate with protein abundances. This trend is robust to controlling for several potentially confounding factors, and is also observed when analyzing protein abundance data for 6 human tissues. In addition, analysis of mRNA abundance data for 32 human tissues shows that the E–R correlation is always less negative, and sometimes non-significant, in secreted proteins. Similar observations were made in *Caenorhabditis elegans* and in *Escherichia coli*, and to a lesser extent in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Arabidopsis thaliana*. Our observations contribute to understand the causes of the E–R anticorrelation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-66

Generalized Mutation-Selection Models by a Diffusion-Free Approach

Ivan Krukov [1,*], A.P. Jason De Koning [1]

[1]Biochemistry and Molecular Biology, University of Calgary, Calgary, Canada

Abstract: Mutation-Selection (M-S) codon-substitution models explain phylogenetic substitutions as the outcome of underlying population-genetic processes. Thereby, they allow inferences about population-level forces like differential allelic fitness to be made from across-species sequence comparisons. M-S models classically rely on diffusion approximation to the behaviour of the underlying Wright-Fisher (W-F) model of population genetics, while ignoring mutation and dominance, and assuming infinite population size and weak selection. Such assumptions may be restrictive in real biological scenarios, where mutation rates can be fast, populations small, and selection strong.

We developed a method for the exact analysis of W-F and other population-genetic/phylogenetic substitution models, which allows us to circumvent diffusion approximations and exactly calculate the rate of phylogenetic substitution while making no additional assumptions over those made by the Wright-Fisher model itself. Since our exact solution does not carry many assumptions, we can model the influence of biological factors that classical M-S can't. By comparing our exact M-S models to their diffusion approximation counterparts, we demonstrate several situations in which classical M-S approaches break down and can lead to incorrect results. In addition, we explore the computational feasibility of the exact M-S models in inference. The exact solutions require more computation than their diffusion counterparts, and results need to be pre-computed across a parameter grid. We investigate biologically relevant parameter ranges, paying special attention to the effect of dominance over the inference of site-heterogeneous fitness profiles.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP3

The impact of native state switching on protein sequence evolution at the residue level

Avital Sharir-Ivry [1], Yu Xia [1,*]

[1]Department of Bioengineering, McGill University, Montreal, Canada

Abstract: A major research focus at the interface between molecular evolution and biophysics is the quantitative elucidation of how protein structure constrains protein evolution at the residue level. Such studies have so far focused on proteins with a single, unique native structure. Many proteins are known to adopt multiple, distinct native structures under different conditions. Such "conformational switches" are expected to exhibit unique evolutionary behavior, yet the impact of such native state switching on protein evolution is not well understood at the residue level.

Here, we performed a proteome-wide analysis of how protein structure impacts sequence evolution at the residue level for protein conformational switches in *S. cerevisiae*. We observed a strong linear relationship between residue evolutionary rate and residue burial for conformational switches. In addition, we found that conformational switches evolve significantly and consistently more slowly at the residue level than proteins with a single native state, even after controlling for degree of residue burial or packing. This is in contrast to flexible and disordered proteins, which are known to evolve more quickly than ordered proteins. Next, we focused on proteins that switch conformations upon molecular binding. We found that interfacial residues in these conformational switches evolve more slowly than interfacial residues in proteins with a single native state, and that the bound conformation is a better predictor for residue evolutionary rate than the unbound conformation.

Our findings suggest that for conformational switches, the necessity to encode multiple distinct native structures under different conditions imposes strong selective constraints on the entire protein, rather than just a few key residues. In addition, the functional, bound conformation imposes stronger selective constraints on these conformational switches than other conformations. Our results provide new insights into the structure-evolution relationship and deeper understanding of the evolutionary design principles of protein conformational switches at the residue level.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-67

Molecular coevolutionary analysis of HIV-1 protein-protein interaction interfaces using phylogenetically independent comparisons

Madara Hetti Arachchilage [1], Helen Piontkivska [1,*]

[1]Biological Sciences, Kent State University, Kent, United States

Abstract: Phylogenetic and stochastic variances must be accounted for in coevolutionary studies, although often this is not a trivial task. Using genomic sequences of Human Immunodeficiency Virus (HIV-1), we examined patterns of coevolution of protein segment pairs between two important players in the HIV-1 life cycle: integrase (IN) and reverse transcriptase (RT). To account for a potential effect of background noise contributed by shared phylogenetic history and/or random interactions, we performed molecular coevolutionary analysis using three separate approaches: randomization, closely-related sister pairs, and moderately divergent sister pairs. Our results showed that insights gained from the analysis of the closely-related sister pairs reduce false positive coevolution signal. On the other hand, analysis of the moderately divergent sister pairs allows us to address potentially false negative coevolution signal derived from stochastic covariation due to lack of divergence in specific protein regions. We identified a set of likely interacting region pairs that exhibited strong coevolution signal consistently in three approaches, thus, making them the most likely candidates for further experimental studies, including development of adjuvant-based treatments and/or novel protein inhibitors that can target functionally and/or structurally important interacting regions in IN and RT.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-68

Natural History of a Catalytically Promiscuous Enzyme Family

Andrew McMillan [1], Benjamin Morse [1], Denis Odokonyero [1], Dat Truong [1], Rebecca Wattam [2], Sing-Hoi Sze [1 3], Margaret Glasner [1,*]

[1]Biochemistry and Biophysics, Texas A&M University, College Station, [2]Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, [3]Computer Science and Engineering, Texas A&M University, College Station, United States

Abstract: Many enzymes are *promiscuous*, catalyzing *non-biological side reactions* in addition to their normal reactions. New metabolic pathways are thought to evolve by recruiting promiscuous enzyme activities. We are evaluating this hypothesis using one of the few natural pathways that is known to use a promiscuous enzyme. This enzyme's ancestral activity was *o*-succinylbenzoate synthesis (OSBS), which is a step in menaquinone synthesis. *N*-succinylamino acid racemization (NSAR) evolved in one branch of the OSBS family, and it is used in a pathway to convert D-amino acids to L-amino acids in some species. By comparing promiscuous NSAR/OSBS enzymes and non-promiscuous OSBS enzymes, we identified two mutations that change the relative specificity of the NSAR and OSBS activities. Neither amino acid directly contacts the substrate, suggesting that catalytic properties of the active site, and not just substrate affinity, affect promiscuity. To determine how metabolic pathways that use NSAR activity evolved, we reconciled the NSAR/OSBS phylogeny and species phylogeny. In addition to a complex pattern of horizontal gene transfer, we discovered a gene duplication in one lineage. To determine how evolution of NSAR/OSBS enzymes correlates with evolution of metabolic pathways that use NSAR, we developed a novel network-based method to analyze potential operons that include an NSAR/OSBS gene. The gene duplication coincides with recruitment of promiscuous NSAR activity into operons for the D- to L-amino acid conversion pathway. We also identified two other operons that suggest NSAR activity is used in bacterial cell wall remodeling. Comparing the phylogenies of genes in these operons is revealing whether NSAR genes were independently recruited into different pathways in different species or whether entire metabolic pathways were horizontally transferred between species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-78

Experimental measurement of ultra high-order epistasis

Lucas Wheeler [1,*], Michael Harms [1], Zachary Sailer [1]

[1]university of oregon, Eugene, United States

Abstract: Epistasis shapes the evolution of new phenotypes by opening and closing potential evolutionary trajectories: the effect of a mutation depends on a previous substitution. This historical contingency could be magnified by interactions between more than two mutations (high-order epistasis), as the effect of a mutation depends on a whole constellation of previous substitutions. We have found that high-order epistasis—consisting of up to sixth-way interactions—is a common feature of experimental genotype-phenotype maps. We used computational approaches to ask whether this high-order epistasis could shape evolutionary outcomes. We found that even low-magnitude high-order epistasis could have strong effects on evolutionary trajectories. This result suggests that the effect of a mutation can depend on up to five previous substitutions. One important question that remains is how far high-order epistasis extends. The barrier to asking this question is the size of previously characterized genotype-phenotype maps, which have at most seven mutations. We therefore set out to measure a genotype-phenotype map containing a larger number of mutations. We engineered a combinatorial phage display library containing all combinations of ten mutations to a 12-mer peptide (2^{10} genotypes). We have done extensive controls and in vitro characterization of peptide binding to validate the assay. We are currently screening the library using two protein partners. Human S100A5 binds to both the wildtype peptide and the peptide containing all 10 mutations. In contrast, human S100A6 binds only to the wildtype peptide. We can simultaneously quantify the enrichment of all members of the screened phage display library using deep sequencing. By screening the library using S100A5, we can study a (nearly) neutral network. By screening the library using S100A6, we can study a network that varies in fitness across its range. We hypothesize that high-order epistasis will continue to 10^{th}-order in both the neutral and variable fitness networks. This experiment will be the largest binary genotype-phenotype map yet characterized, revealing how the magnitude of epistasis changes with increasing order of interaction.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-106

The impact of gene remodelling events on fungal evolution.

Robert Leigh [1,*], David Fitzpatrick [1], James McInerney [2]

[1]Department of Biology, Maynooth University, Maynooth, Ireland, [2]Faculty of Biology, Medicine, and Health, The University of Manchester, Manchester, United Kingdom

Abstract: Gene remodelling is the biological process of rearranging sequence nucleotides that may result in the translation of new gene products. Remodelling may occur due to point mutations, chromosomal rearrangement, or mosaic events. Mosaicism describes a reportedly rare remodelling event where two or more protein coding genes may undergo a fusion event to form a single protein-coding gene, or undergo a fission event leading to one or more independent protein coding genes. By assessing mosaicism in protein association networks, gene neofunctionalisation and distribution patterns we can hypothesize evolutionary pressures behind these events and how these events may have conferred host advantages. Fusions are reported to be more common than fissions due to their relative mechanistic ease. Fusions may arise when start codon deletion occurs on one gene and stop codon deletion occurs on another, allowing them to be read as one continuous ORF. Comparatively, fission arises due to the breakage of an ORF where each segment requires the gain of a stop codon, gain of a promoter region and the appropriation of a start codon. We employed a number of bioinformatic techniques to assess these events across 108 sequenced fungal taxa and report a large number of mosaic events despite strict filtration techniques applied. By utilizing a filtered protein-protein interaction of our mosaic events we report high rates of gene remodelling amongst all of these species. We are currently investigating the biological significance of these gene families and assessing if they have a role in phenotype, niche and pathogenicity specification.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-103

Rhodopsin function elucidates visual adaptation to dim light and highlights contrasting sensory specializations in bats

Eduardo de Almeida Gutierrez [1,*], Gianni Castiglione [2], James Morrow [2], Belinda Chang [1]

[1]Ecology and Evolutionary Biology, [2]Cell and Systems Biology, University of Toronto, Toronto, Canada

Abstract: The nocturnal environment poses strong selective constraints on how animals perceive and interact with their surroundings. Given their long history of nocturnality, bats are a fascinating system in which to study how sensory systems evolved in light limiting conditions. Although most bats are able to navigate in the dark through sophisticated echolocation calls, they also have a developed visual system that operates well in dim light. This ability is mediated at the molecular level by the visual pigment rhodopsin, a transmembrane photosensitive protein expressed in the outer segment of rods. Functional properties of rhodopsin contribute to maximizing photon capture, influencing visual system sensitivity. Interestingly, contrasting visual abilities observed in bats are apparently associated with different echolocation capabilities. To investigate this association, we expressed and compared rhodopsin function in three bat species with distinct sensory specializations. Bat rhodopsin genes were synthesized and heterologously expressed *in vitro*. Purified rhodopsin samples were then functionally assayed through absorbance spectroscopy and fluorescence-based assays to assess spectral and kinetic differences between the three bats and a bovine control. While maximum absorbance showed little variation, rhodopsin kinetics varied significantly among bats, suggesting that differences in dim-light photosensitivity are associated with distinct echolocation capabilities. Moreover, rhodopsin kinetics was also significantly different between all bats and bovine. To further investigate these pronounced functional differences, we mutated key amino acid sites that vary between bats and bovine, and functionally assayed rhodopsin mutants. Here we report our results and discuss the molecular mechanisms underlying increased photosensitivity in bats.

Expanded summary*: Photic-limiting environments act as strong selective constraints on the evolution of sensory systems, influencing how animals perceive, interact and respond to their natural surroundings. The visual system is particularly modified in organisms that have evolved in dim-light conditions, exhibiting either a suite of anatomical, morphological and molecular modifications to maximize photon capture, or a systemic degeneration, frequently accompanied by extensive loss of visual gene function and greater specialization of a light-independent sensory modality.

Having long occupied the nocturnal environment, the order Chiroptera represents one of the most ecologically diverse mammalian lineages and exhibit an array of sensory adaptations to dim light. Bats are most notable for their echolocation abilities, a key sensory innovation that allows bats to navigate and forage in complete darkness. On the other hand, many bats also have a developed visual system that operates well in dim light and display anatomical and morphological features consistent with dim-light adaptation, such as an enlarged cornea and a rod-dominated retina. Interestingly, contrasting visual abilities are observed in bats presumably as a result of a sensory trade-off associated with different echolocation capabilities, as observed by visual gene pseudogenization in some echolocating lineages.

Dim-light vision is mediated by the visual pigment rhodopsin, a seven-transmembrane protein expressed in the outer segment of rod cells. Rhodopsin comprises a photosensitive complex bound to a vitamin A-derived chromophore that isomerizes upon stimulation by light, leading to activation of the visual transduction cascade and further interpretation of the visual signal in the brain. Functional properties of rhodopsin are determined by several key amino acid sites that influence spectral tuning and protein kinetics. Substitutions at these sites contribute to maximizing photon capture and influence visual system sensitivity, resulting in adaptation to varying light environments. The molecular underpinnings underlying dim-light adaptation in bats have not been extensively studied. To investigate the contrasting visual abilities within Chiroptera, we expressed and compared rhodopsin function of bats with distinct sensory specializations: a non-echolocating (NE) and presumably more visual species, and two echolocating species with distinct abilities, a low-duty cycle (LDC) and a high-duty cycle (HDC) bat. We hypothesized that rhodopsin function would differ among these species and that differences would be consistent with the degree of reliance on vision relative to echolocation. Bat rhodopsin sequences obtained from publicly available genomes were synthesized, heterogously expressed *in vitro* and purified using an immunoaffinity protocol. Purified rhodopsin samples were then functionally assayed through absorbance spectroscopy and fluorescence-based assays to assess spectral and kinetic differences, respectively, among the three species. While maximum absorbance showed little variation, rhodopsin kinetics varied significantly among bats. Kinetic differences were consistent with the

type of echolocation, suggesting a trade-off between dim-light photosensitivity and echolocation capability. Moreover, rhodopsin kinetics of all bats were also remarkably different when compared to bovine rhodopsin, a model for terrestrial mammalian systems. To further investigate this pronounced functional difference, we mutated key amino sites that vary between bats and bovine, and functionally assayed rhodopsin mutants. Here we report our preliminary results and discuss the possible molecular mechanisms mediating increased photosensitivity and rhodopsin adaptation in bats, which can potentially be applied to understand dim-light vision evolution in other mammals.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-64

Co-evolution of zinc finger C2H2 proteins and endogenous retroelements in human and mouse

Marjan Barazandeh [1,*], Timothy R. Hughes [1]

[1]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada

Abstract: C2H2 zinc finger proteins (ZFPs) are the largest group of transcription factors (TFs) in human, yet the role of many of them are unknown. C2H2 ZFPs are rapidly evolving and in vertebrates often contain KRAB (a potent transcriptional repressor domain). A recent hypothesis proposes that KRAB ZFPs may evolve to silence endogenous retroelements (EREs). EREs make up half of the human genome, expand throughout the genome via retrotransposition and may cause diseases by invading the genome regulatory regions. Therefore, studying the natural mechanisms of silencing them is of great importance. Recent studies indicate that a majority of KRAB ZFPs in human bind specific classes of EREs, thus may inhibit their further expansion across the genome. However, the sequence preferences of KRAB ZFPs are largely unknown in other species. Therefore, we used ChIP-seq to detect the binding sites of the orthologous KRAB ZFPs in human and mouse. We expressed both human and mouse genes in human cells. Interestingly, we observed nearly identical motifs for orthologous ZFPs and enrichment for the same ERE instances. However, the non-ERE binding sites are slightly different between species. Moreover, similar protein-protein interactions were observed between the orthologous ZFPs. This suggests that ERE binding is conserved between species, but ZFPs might have also acquired independent host functions at non-ERE binding sites. We are also conducting the same experiments in mouse cell lines. This study extends our knowledge about the evolution of a major class of TFs and mechanisms behind the transcriptional regulation of complex genomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-105

Detecting protein-protein interactions between Hydractinia allorecognition proteins with an ELISA-based assay

Aidan Huene [1,*], Matthew Nicotra [1]

[1]University of Pittsburgh, Pittsburgh, United States

Abstract: Allorecognition is the ability to distinguish between self and genetically distinct conspecifics. In the hydroid *Hydractinia*, allorecognition determines whether colonies aggressively compete for space or fuse to form a single colony when they encounter each other as they grow. Studies in inbred lines of *Hydractinia* indicate that allorecognition is controlled by the Allorecognition Complex (ARC), a genomic region containing at least two allorecognition genes. These genes (*Alr1*/*Alr2*) are highly polymorphic, encode transmembrane proteins and have IgSF-like ectodomains. Using an *in vitro* assay, we have examined four Alr1 and six Alr2 alleles and determined each selectively binds across opposing cell membranes to itself or nearly identical isoforms. In nature, more than 180 allelic isoforms of Alr2 have been documented, and Alr1 is expected to be similarly diverse. If most of these isoforms are capable of isoform-specific homophilic binding and are required for colonies to fuse, it would provide a mechanistic explanation for the low rates of fusion (~2%) observed in nature. A related question is how hundreds of alleles with unique homophilic binding specificities could evolve. By testing isoforms in a pairwise format, the binding specificity locales can be identified. Here we describe the development of an ELISA-like assay to efficiently screen the Alr binding properties and identify the regions of the ectodomains that contribute to binding specificity. We anticipate the results of these studies will reveal the structural mechanisms that underlie homophilic binding in one of the most polymorphic genetic systems identified to date.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-74

Tracing the evolution of protein-RNA and protein-protein interactions across animal double-stranded RNA-binding proteins (DRBs)

Kelsey Aadland [1,*], Bryan Kolaczkowski [1]

[1]Microbiology and Cell Science, University of Florida, Gainesville, United States

Abstract: Understanding how proteins evolve affinity for new ligands and interactions with new partner proteins is central to molecular evolutionary biology and critical for linking changes in molecular function to changes in cellular processes. Ancestral sequence reconstruction has allowed researchers to rigorously dissect the evolutionary and structural mechanisms driving individual changes in ligand preference at particular nodes on a protein family phylogeny. However, determining how ligand preference evolves over longer timescales has proven more difficult, and the evolution of protein-protein interactions has rarely been investigated mechanistically. Here we present the first study combining ancestral sequence reconstruction, structural modeling and ligand affinity prediction to explicitly examine the structural basis for changes in ligand affinity and protein-protein interactions across the entire evolutionary history of a protein family. We apply this novel approach to examine the long-term evolution of animal double-stranded RNA-binding proteins (DRBs), a family that contributes to RNA interference by directly mediating the interaction between Dicer and its RNA target. We show how individual domains within the DRB protein differentiated into RNA-binding and protein-binding domains, characterize the extent to which molecular-functional 'trade offs' may have constrained potential evolutionary trajectories and evaluate how co-evolution between DRBs and their Dicer partners affected long-term DRB-Dicer interactions. Our work demonstrates the potential for high-throughput ancestral protein resurrection to shed new light on how protein families evolve functional diversity across very large phylogenies, informing a deeper understanding of how molecular evolutionary processes work.

Expanded summary*: Central to the study of molecular and evolutionary biology lies protein functional evolution and how proteins evolve affinity for new ligands and interactions over time. While studies have been done on protein families and the change in their ligand binding affinities, these studies have only been done over short evolutionary timespans for these proteins. In my research, I've used a unique and explicitly mechanistic method for characterizing the change in protein functional evolution over the entire evolutionary history of a protein family.

In this study, I employ Ancestral Sequence Reconstruction (ASR), structural protein modeling, and molecular dynamics simulations to predict changes in ligand binding affinity and protein-protein interactions across long evolutionary distances. I utilize this approach in my examination of the functional evolution of animal double-stranded RNA binding proteins (DRBs), which are involved in RNA interference by aiding the interaction between Dicer and its RNA target.

The evolution of DRBs is particularly interesting because of their tandem repeat double-stranded RNA-binding motif (dsrm) domains, which over time are believed to have evolved multiple functions. Recent evidence has led researchers to believe that two out of the three tandem repeat domains are involved in the double-stranded RNA-DRB interaction, while the third tandem repeat domain is involved in the DRB-Dicer protein-protein interaction.

This project provides a unique opportunity to not only study the functional evolution of proteins over long phylogenetic distances, but also to characterize how evolutionary trajectories might have changed due to molecular-functional 'trade-offs' in regards to RNA-binding and protein-binding domains. Along with this, it allows for the evaluation of patterns of co-evolution between the DRB and dicer protein families and how this affected DRB-dicer interactions over long periods of time.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-81

Are nonsynonymous transversions more deleterious than nonsynonymous transitions?

Zhengting Zou [1,*], Jianzhi Zhang [2]

[1]Department of Computational Medicine and Bioinformatics, [2]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, United States

Abstract: Transitions typically have higher substitution rates than transversions in coding sequence evolution. This phenomenon has several potential causes. First, the mutation rate is higher for transitions than transversions. Second, transitional mutations are more likely than transversional mutations to be synonymous and hence have a higher rate of fixation. Third, it has been suggested that transitional nonsynonymous mutations are more likely than transversional nonsynonymous mutations to conserve amino acid physicochemical properties and so have a higher rate of acceptance. This third possibility was recently found untrue because of no detectable difference in fitness effects between transitional and transversional nonsynonymous mutations in large mutagenesis experiments. However, because laboratory measures of fitness effects have limited sensitivity, we used evolutionary data to revisit this issue. We modified an existing codon model of sequence evolution by adding a new parameter η, which is the ratio between the probability of acceptance of a transitional nonsynonymous mutation and that of a transversional nonsynonymous mutation. Using a likelihood estimator of η, we examined genome-wide concatenated alignments of coding sequences from many species pairs across the tree of life. Surprisingly, η varies widely from smaller than 1 to greater than 1. Thus, in some species, transitional nonsynonymous mutations are more deleterious than transversional nonsynonymous mutations, but the opposite is true in some other species. Our extensive searches reveal that this diversity arises from variable amino acid exchangeabilities across the tree of life. The causes and ramifications of among-species variations of such fundamental parameters of evolution await further exploration.

Expanded summary*: There have been conflicting conclusions about whether nonsynonymous transitions are more or less deleterious than nonsynonymous transversions (Zhang, 2000; Stoltzfus and Norris, 2016). In this study, we attempted to resolve this debate using widely sampled comparative genomic datasets and molecular evolutionary analysis. It turns out that the answer varies depending on the species concerned. After an extensive search, we found that the primary cause is an among-species variation in amino acid exchangeabilities among species. This is biologically unexpected and intriguing, because it is widely assumed that a single largely constant matrix of amino acid exchangeabilities applies in all species. Note that here exchangeability refers to the relative probability of fixation of an amino acid change. That is, the common belief that certain amino acid changes are always readily acceptable (on average across a genome) and certain amino acid changes are always difficult to accept no matter which species is concerned is wrong.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-86

Present performances of the old enzyme deoxyuridine triphosphatase

Hideaki Moriyama [1,*]

[1]School Of Biological Sciences, University of Nebraska, Lincoln, United States

Abstract: deoxyuridine triphosphatase (dUTPase) is an "old enzyme," which appeared relatively soon after the emergence of life. dUTPase genes have now evolved to enable dUTPase to perform various roles, including different enzyme activities and as a scaffold in viral capsids, while maintaining signature features, such as high heat stability. However, we still have to address the question why does dUTPase found in *Chlorella* virus have a high optimal temperature of >353 K when its hosts live at approximately 298 K?
Two forms of dUTPase from chlorella viruses IL-3A and PBCV-1 differ in their optimal temperature activity by 15 K; the PBCV-1 enzyme has the higher temperature activity 353 K. Only 9 of 141 amino acids differ between the two dUTPases. Glu81 and Thr84 are part of the active site in the IL-3A enzyme, whereas the corresponding active site residues are Ser81 and Arg84 in the PBCV-1 enzyme. We constructed a mutated dUTPase, Mu-22, by replacing the two amino acids in IL-3A with the two PBCV-1 amino acids. These two amino acid substitutions increased the temperature optimum of the IL-3A enzyme to 360 K. To elucidate the molecular mechanism(s) behind this large shift in temperature optimum, we solved the 3D-crystal structures of the IL-3A and Mu-22 dUTPases. Both dUTPases were homotrimers and bound three molecules of the inhibitor dUDP. The active site structures resembled human dUTPase. In Mu-22, the active site lost rigidity and the cavity at the center of the trimer increased in volume by 20%. X-ray small angle scattering experiments determined that the radius of gyration of IL-3A and Mu-22 dUTPases were 22 Å and 20 Å, respectively. The molecular size of both enzymes decreased in the presence of dUDP. The structural denaturing temperatures measured by circular dichroism were 338 K and 335K for IL-3A and Mu-22 enzymes, respectively. The shift in temperature optima for enzyme activity was interpreted as a trade-off between enzyme stability and chemical reactivity.
Nevertheless, the energy cost seemed too expensive for the chemical reaction. It was because the enzyme cleaves one phosphor dieter bonding only. We searched other factors to determine the kinetic performance of the enzyme by focusing the behaviors of water molecule in the active site. Our exhaustive structural mining indicated that the affinity in the "on" mode is governed by a single water molecule held by hydroxyl and imidazole groups. When the enzyme has the SxxxW motif, the affinity reaches to the maximum, which find in half of the plant.
Zhang, Y., Moriyama, H., Homma, K., and Van Etten, J. L. (2005) Chlorella virus-encoded deoxyuridine triphosphatases exhibit different temperature optima. J Virol 79, 9945-9953.
Noriko Inoguchi; Kittichai Chaiseeda; Mamoru Yamanishi; Moon Ki Kim; Yunho Jang; Mamta Bajaj; Catherine P. Chia; Donald Becker; Hideaki Moriyama. (2015). Structural insights into the mechanism defining substrate affinity in Arabidopsis thaliana dUTPase: The role of tryptophan 93 in ligand orientation. BMC Research Notes. 8:784.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-87

Directed Evolution of human L-kynureninase for the Development of an Enzyme Acting as Immune Checkpoint Inhibitor Agent

Christos S. Karamitros [1,*], John Blazeck [1], Kyle Ford [1], Kendra Garrison [1], Todd Triplett [2], Everett Stone [2], George Georgiou [1]

[1]Chemical Engineering, [2]Molecular Biosciences, UT Austin, Austin, United States

Abstract: L-Kynureninases (L-Kynases; E.C. 3.7.1.3) are PLP-dependent enzymes that belong to the aminotransferase superfamily of enzymes and catalyze the conversion of L-Kynurenine and 3'-OH-L-Kynurenine to anthranilic acid and L-alanine or 3'-OH-anthranilic acid and L-alanine, respectively. It has been shown that elevated levels of L-Kyn can inhibit and suppress the adaptive immune system, thereby promoting cancer escape and development. ***Our hypothesis*** is that serum or tumor microenvironment depletion of the increased levels of L-Kyn by using an enzyme capable of degrading it into nonfunctioning agents (anthranilic acid and L-alanine) would have a significantly positive impact on the restoration of the adaptive immune system, thus promoting its anti-tumor activity. Therefore, we have been motivated to **engineer an L-Kynase of human origin** (hKynase) aiming at the development of this enzyme as an immune checkpoint inhibitor agent. Given the fact that non-human enzymes elicit severe immunogenic responses during the treatment of humans, the use of a human enzyme would eliminate the risk of developing immune reactions.

L-Kynases are able to accept both L-Kyn and 3'-OH-L-Kyn as substrates, yet with distinct activities and specificities. Bacterial enzymes are characterized by high activity towards L-Kyn (k_{cat} ~ 5-10 s^{-1} and K_M ~ 50 μM), while mammalian L-Kynases including hKynase are more efficient against 3'-OH-L-Kyn (k_{cat} ~ of 5-10 s^{-1} and K_M ~ of 50 μM) hydrolyzing only poorly L-Kyn. Amino acid alignments between bacterial and mammalian L-Kynases reveal low degree of identity (20-30%), yet the tertiary structures of these enzymes are remarkably similar. In addition, combination of phylogenetic and structural analyses of bacterial and mammalian enzymes helped us to identify key-residues that confer substrate specificity. Briefly, the bacterial enzymes that hydrolyze L-Kyn with high activity are characterized by the triad "Trp-Gly-Thr" present in the active site. In contrast, the mammalian enzymes have "His-Ser-Asn" at the respective positions in the active site. Extensive mutagenesis and directed evolution experiments by applying variety of methods coupled with a very efficient and stringent genetic selection approach allowed us to identify the role of the active site residues mentioned above, as well as the role of distal mutations in hKynase. The applied methods include epPCR, site-directed mutagenesis, DNA shuffling, Rosetta-based computational analysis, domain insertion/deletion and generation of bacterial/hKynase chimeras. The genetic selection approach relies on the use an L-Trp auxotrophic *E.coli* strain whose genetic defect is complemented by the activity of hKynase. Cells expressing hKynase variants with improved activity and stability crowd out the rest of the cells and dominate the population in the culture after successive rounds of sub-culturing of the initial mutant library. In the last screening step, we characterize the enzymatic activity of individual clones on 96-well micro titer plates by employing an enzymatic assay that monitors the rate of the L-Kyn depletion at 365 nm. Taken together, the results we have gathered so far from the biochemical and kinetic analyses of numerous hKynase variants, highlight the importance and the specific roles of certain residue networks that confer greatly improved activity and specificity of hKynase towards L-Kyn, as well as considerably improved serum and conformational thermodynamic stability.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP11

Tuning the course of evolution on the biophysical fitness landscape of an RNA virus using droplet microfluidics

Adrian Serohijos*

Abstract: Tuning the course of viral evolution on a biophysical fitness landscape using droplet microfluidics

Adrian Serohijos[1,3], Assaf Rotem[2], James M. Pipas[5], Christiane E. Wobus[4], Andrew B. Feldman[6], David A. Weitz[2], and Eugene I. Shakhnovich[3]

[1]Département de Biochimie et Centre Robert-Cedergren en Bioinformatique et Génomique, Université de Montréal, Quebec, Canada;
[2] School of Engineering and Applied Sciences and Department of Physics, Harvard University, 9 Oxford Street, Cambridge, MA 02138, USA
[3] Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA
[4] Department of Microbiology and Immunology, University of Michigan Medical School, 1150 West Medical Center Drive, Ann Arbor, MI 48109, USA
[5] Department of Biological Sciences, University of Pittsburgh, 4249 Fifth Avenue, Pittsburgh, PA 15260, USA.
[6] Department of Emergency Medicine, Johns Hopkins Medicine, 5801 Smith Avenue, Suite 3220, Davis Building, Baltimore, MD, 21209, USA

Evolution is a unifying theme in the urgent medical and public health problems we face today including cancer, the rise of antibiotic resistance, and the spread of pathogens. But the ability to predict evolution remains a major challenge because it requires bridging several scales of biological organization. Potential evolutionary pathways are determined by the "fitness landscape" (the genotype-phenotype relationship), but how this landscape is explored depends on microbial population dynamics.
I will describe our recent work where we showed that the fitness landscape of norovirus escaping a neutralizing antibody can be projected onto two traits, the capsid folding stability and its binding affinity to the antibody. We then developed a theory based on protein biophysics and population genetics to predict how the fitness landscape might be explored. Using a droplet-based microfluidics "Evolution Chip", we propagated millions of independent viral sub-populations, and showed that by tuning viral population size per drop, we could control the direction of viral evolution. Additionally, I will describe how this combined framework of biophysics and evolutionary biology also applies to bacterial evolution due to horizontal gene transfer. Altogether, these stories demonstrate the broad applicability of the techniques and concepts from protein engineering to fundamental problems in evolution and genetics.
Keywords: folding stability, binding affinity, viral evolution, RNA virus, population genetics, Wright's Shifting Balance Theory, neutralizing antibody

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP12

Evolution of a conformational ensemble

Jessica Siltberg-Liberles [1 2,*], Helena G. Dos Santos [1]

[1]Biological Sciences, [2]Biomolecular Sciences Institute, Florida International University, Miami, United States

Abstract: We recently found clade-specific variation in predicted intrinsic disorder for 543 tyrosine kinases across 17 paralogous clades. To explicitly study the dynamics from actual 3D models of protein structure, here we present a large-scale study based on 3D models for the same 543 sequences. For every sequence, a conformational ensemble of 7 representative conformations was modelled and analyzed with normal modes analysis. Further, 25 ns molecular dynamic simulations were performed on one conformation for the 17 human representatives, and for four sister clades including 7 vertebrates (*Danio rerio, Gallus gallus, Monodelphis domestica, Sarcophilus harrisii, Bos taurus, Mus musculus,* and *Homo sapiens*). From these analyses we infer (i) the different conformations are accessible to all proteins but some are preferred for different clades, (ii) extensive variation in the amount of intrinsic dynamics persists within and between different paralogous clades, (iii) clade-specific shifts in stability, partly due to shifts in the electrostatic contribution to the free energy, are apparent and conformation-independent, and (iv) large shifts in electrostatic potential surfaces for different clades regardless of conformation. Lastly, one of the paralogous clades has become significantly more dynamic and divergent in multiple measures and appears to have neofunctionalized to perform its function via a unique mechanism of binding to and stabilizing its paralogs, as also confirmed by previous experimental studies. Overall, the divergence of protein dynamics after gene duplication is further supported by these results, implying the importance of altered dynamics on functional divergence.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP4

Using structural conservation to evaluate biological realism in models of protein evolution

Claudia Weber [1,*], David Liberles [1]

[1]Temple University, Philadelphia, United States

Abstract: Phylogenetic models are important for characterizing how proteins evolve. To provide a meaningful reflection of the evolutionary process, they should ideally capture selective constraints acting on coding sequences. Statistical fit to observed sequences, which assumes that sites evolve independently, is typically used to infer how well the models describe the data. But are models that fit empirical data better also more biologically realistic? How can we determine which features are likely to improve models without requiring a full-scale implementation in a phylogenetic inference framework?

To answer these questions, we performed forward simulations and examined how well different models preserve protein structure at varying levels of divergence. We find that an empirical model (LG08) that incorporates amino acid exchangeabilities results in less deviation in secondary and tertiary structure than a neutral model that considers only the structure of the genetic code. Because LG08 does not account for site-specific preferences, one might expect that a more complex model is more realistic. Indeed, scaling substitution rates according to observed patterns of evolution (LG08+4dG) slows structural deviation, as well as providing improved statistical fit. This is not due to a greater number of conserved slow-evolving residues, as shuffling rate factors randomly between sites results in a decay curve resembling LG08. We additionally examined conservation of residue solvent exposure.

Nevertheless, a model that assumes that all sites are independent and share a common set of amino acid frequencies towards which they evolve is somewhat dissatisfying. We therefore also tested a selection scheme that accounts for contact affinities (Miyazawa & Jernigan, 1985; Williams, 2006) between residues and scores mutants on the stability of the native structure. While this simple model performs better than both the empirical and neutral models in terms of tertiary structure at short time scales, structural deviation actually accelerates thereafter, and is overall worse for secondary structure. This observation is perhaps consistent with the lack of constraints imposed on features besides contacts between residues. However, the initial advantage of the structurally constrained simulation over conventional methods indicates that there is room for improvement with more elaborate biochemically and structurally informed models. More generally the agreement between statistical fit and structural conservation for models of a similar class suggests that forward simulations may be useful to assess which selection criteria promise to provide a more accurate description of protein evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-85

A Novel Method to Identify Proteins with Correlated Rates of Evolution

Andrew Webb [1,*], Jody Hey [1]

[1]Department of Biology, Temple University, Philadelphia, United States

Abstract: Here we describe a new methodology capable of assessing correlation in large numbers of proteins, which demonstrates how estimated branch lengths fall naturally into a contingency table framework which can then be easily used to infer relative rates of evolution. To investigate the capacity of our method to identify these correlations we used a combination of simulated datasets and mammalian genomic alignments. Our findings demonstrate how this framework allows for rapid identification of proteins that are potentially co-evolving. We also explore how to extend our framework to allow for missing data among the phylogeny and the potential biological causes of our observed correlated proteins.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POB-420

REPEATED EVOLUTION OF THE ENZYME ACTIVITY OF HOMOSPERMIDINE SYNTHASE IN PLANT SECONDARY METABOLISM

Arunraj Saranya Prakashrao [1,*], Elisabeth Kaltenegger [1], Dietrich Ober [1]

[1]Kiel University, Kiel, Germany

Poster: Gene duplication followed by functionalization plays a major role in the origin and distribution of enzymes in secondary metabolism. Homospermidine synthase (HSS) the enzyme catalyses the first and critical step in pyrrolizidine alkaloid (PA) biosynthesis evolved from deoxyhypusine synthase (DHS). To understand the mechanism that shaped the evolution of HSS activity, we identified HSS and DHS sequences from different plant species of Morning glory family. Phylogenetic analysis showed that a single gene duplication event gave rise to extant HSS sequences in many species. The duplicated HSS paralogs underwent varying selection pressures throughout their evolution including purifying selection, neutral, and positive Darwinian selection. Reconstructing the ancestors of extant HSS sequences, reveals identical amino acid replacements in the HSS after the gene duplication. Functional characterization of ancestral and extant HSS sequences signals a pattern of repeated evolution of HSS activity in different species of Morning glory.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POB-432

INVESTIGATING ACTIVE SITE DYNAMICS OF APICOMPLEXANS MALATE DEHYDROGENASE BY NUCLEAR MAGNETIC RESONANCE

Miriam Hood [1,*], Jacob Wirth [1], Michelle Fry [1], Jeffrey Boucher [1], Brian Beckett [1], Douglas Theobald [1]

[1]Brandeis University, Waltham, United States

Poster: A variety of human diseases including Malaria are caused by unicellular eukaryotes of the genus Apicomplexa, which have evolved highly specific lactate dehydrogenase (LDH) from malate dehydrogenase (MDH). The importance of LDH to the parasitic life cycles of modern Apicomplexans makes it a major drug target. The presence of a potential promiscuous intermediate in the evolutionary path means the Apicomplexan LDHs and MDHs are an excellent model for studying the role of promiscuous intermediates in the evolution of specificity. The two enzymes share structural and mechanistic similarities but differ in substrate specificity. A six amino acid insert is responsible for the development of pyruvate activity in LDHs, which shifts the key catalytic residue from Arg102 to Trp107f. There are two hypothesized conformations for the active site of the bifunctional ancestor MDH with the six amino acid insert. An LDH conformation of the ancestor in the presence of lactate is seen in X-ray crystallography, but an MDH conformation is not seen. Heteronuclear single quantum coherence nuclear magnetic resonance (HSQC-NMR) was attempted to visualize the enzyme in the presence of oxaloacetate but was unsuccessful due to the size of the enzyme. Disruptive mutations were used to create a stable dimer, small enough to produce a well resolved HSQC NMR spectrum. Purification and steady state kinetic assays of the dimer with both LDH and MDH substrates confirm the bifunctionality of the enzyme. HSQC NMR of the dimer with oxaloacetate and pyruvate is well resolved, but does not show significant peak shifts. Isothermal calorimetry (ITC) confirms low micromolar K_Ds for the dimer with NADH and $NADH_4$ and preliminary differential scanning calorimetry (DSC) indicates a T_m of 67°C.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POB-410

GENETIC BASIS OF BRAIN SIZE EVOLUTION IN CETACEANS: INSIGHTS FROM ADAPTIVE EVOLUTION OF SEVEN PRIMARY MICROCEPHALY (MCPH) GENES

Shixia Xu*, Xiaohui Sun [1], Xu Niu [1], Zepeng Zhang [1], Ran Tian [1], Wenhua Ren [1], Kaiya Zhou [1], Guang Yang [1]

[1]Nanjing Normal University, nanjing, China

Poster: Cetacean brain size expansion is an enigmatic event in mammalian evolution, yet investigation of its genetic basis has begun only recently. Here, all exons of the seven primary microcephaly (*MCPH*) genes that play key roles in size regulation during brain development were examined in representative cetacean lineages. Extensive positive selection was identified in four of six intact MCPH genes: *WDR62*, *CDK5RAP2*, *CEP152*, and *ASPM*. Specially, positive selection at *CDK5RAP2* and *ASPM* were examined along lineages of odontocetes with increased encephalization quotients (EQ) and mysticetes with reduced EQ but at *WDR62* only found along odontocete lineages, which is well matched with cetacean complex brain size evolution. Very interestingly, a positive association between evolutionary rate (ω) and EQ was identified for *CDK5RAP2* and *ASPM*, suggesting that these two genes may have contributed to EQ expansion in cetaceans. This suggestion was further indicated by our preliminary function test that *ASPM* might be mainly linked to evolutionary increases in EQ. Most strikingly, our results suggested that cetaceans evolved large brains to manage complex social systems, consisting with the 'social brain hypothesis', as evolutionary rate of *ASPM* and *CDK5RAP2* were significantly related to mean group size (as one measure of social complexity).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP10

High-order epistasis is inevitable in molecular systems

Zachary Sailer [1,*], Michael Harms [1 2]

[1]Chemistry and Biochemistry, [2]Institute of Molecular Biology, University of Oregon, Eugene, United States

Abstract: High-order epistasis (or multi-way interactions between loci) is present across a wide variety of genotype-phenotype maps. Such interactions can profoundly shape evolutionary outcomes, as the effect of a new mutation is determined by interactions with a whole collection of previous substitutions. The origin of this epistasis is not well understood, but its ubiquity points to a universal process, rather than something specific to each genotype-phenotype map. One shared feature of molecular systems is the ability to take multiple configurations in response to stimuli—whether an allosteric protein changing structure or a transcriptional network changing state. We set out to probe whether such systems naturally give rise to high-order epistasis. We used a simple protein lattice model, which allows an amino acid sequence to populate an "ensemble" of possible structures. The probability of each structure is determined by pairwise, physical interactions between amino acids in the sequence. Mutations to the sequence alter the probabilities of individual structures in the ensemble. Using this model, we observed up to 10-way epistasis between mutations, even though the model was built from purely pairwise interactions. The high-order epistasis arises because mutations alter the relative probabilities of multiple structures at the same time. The high-order epistasis can be removed if we constrain the model to take a single structure. If we simulate evolution through lattice model spaces with and without high-order epistasis, we observe radically different trajectories. This points to a profound link between the physical chemistry of molecular systems and the accessibility of evolutionary trajectories.

Expanded summary*: High-order epistasis – or multi-way interactions between loci – has been observed in many genotype-phenotype maps. These interactions lead to profound evolutionary contingency, where the effect of a future mutation depends on a large collection of past substitutions. While pairwise epistasis can often be interpreted in simple, mechanistic terms, the origin and interpretation of high-order epistasis is much less clear. What does a five-way interaction mean? Where does it come from? Are such interactions a general feature of genotype-phenotype maps, or do they only rarely occur?

To understand the origins of high-order epistasis, we took a computational approach. We used simple, physics-based "lattice" models of proteins to simulate large genotype-phenotype maps. Each genotype is a linear sequence of amino acids. This then folds to many different possible structures, with the probability of each structure determined by pairwise energetic interactions between amino acids. The phenotype is the relative probability of a single functional structure, relative to all other structures. This captures a key feature of real proteins: each sequence can take on multiple structures, but the function of the protein is usually determined by a single native structure.

Although the lattice models are built from simple, pairwise interactions, the map itself exhibited extensive high-order epistasis. Because these computational models have known "physics," we could then understand the origin of this epistasis, shedding light on the likely origins of high-order epistasis in real biomolecular systems.

High-order epistasis arises because mutations affect the relative probabilities of all possible structures in a nonlinear fashion. As mutations accumulate, the energy of each structure changes relative to other structures in the "ensemble" of possible structures. This means that the effect of any mutation depends strongly on other mutations that have occurred before it. We showed this directly by simulating proteins that could only take a single structure. Such models do not exhibit high-order epistasis.

Further, we found that high-order epistasis dramatically altered evolutionary trajectories in these genotype-phenotype maps. We ran evolutionary simulations in genotype-phenotype maps both with and without the multi-state structural ensemble. Under the same evolutionary conditions, we see radically different evolutionary trajectories.

This previously unrecognized source of epistasis suggests a profound link between the underlying chemistry of biological systems and the evolutionary process. High-order epistasis can arise from a very simple model, with the only requirement being that each genotype can take multiple states. All molecular biological systems—from individual proteins to regulatory networks—take multiple states, suggesting that high-order epistasis is a universal feature of molecular systems. Because high-order epistasis profoundly alters evolutionary trajectories, our work reveals that the chemistry of biomolecular systems inevitably leads to extensive, multi-mutation contingency in genotype-phenotype maps.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-79

The Apicomplexa-specific glucosamine-6-phosphate N-acetyltransferase gene family encodes an enzyme key for glycan synthesis

Marta Cova [1], Sara Artigas-Jerónimo [1], Aida González [1], Borja López-Gutiérrez [1], Giulia Bandini [1], Yves Van de Peer [2], Steven Maere [2], Luis Izquierdo [1], Lorenzo Carretero-Paulet [2,*]

[1]ISGlobal, Barcelona Ctr. Int. Health Res. (CRESIB), Hospital Clínic – Universitat de Barcelona, Barcelona, Spain, [2]Center for Plant Systems Biology, VIB-UGent, Ghent (Belgium), Belgium

Abstract: The apicomplexan form a phylum of obligate parasitic protozoa, which include species of great clinical and veterinary importance. As every eukaryotic cell, they synthesize glycoconjugates for their protection and to interact and respond to changes in their environment. Among them, Glycosylphosphatidylinositol (GPI) anchors are one of the major carbohydrate modifications described in Apicomplexa. However, despite its importance for their survival and pathogenicity, some of the enzymatic steps involved in GPI biosynthesis remain elusive. In particular, Glucosamine-phosphate *N*-acetyltransferase (GNA) activity, involved in the generation of the UDP-N-acetylglucosamine (UDP-GlcNAc) precursor needed to feed GPI anchor biosynthesis and also *N*-glycosylation, had not been identified yet in any apicomplexan species. We scanned the genomes of *Plasmodium falciparum*, and representatives from six additional main lineages of the phylum, for proteins containing the Gcn5-related *N*-acetyltransferases (GNAT) domain. One of them, designated as *Pf*GNA, was able to rescue the growth of a *Saccharomyces cerevisiae* temperature sensitive GNA mutant. Sequence, phylogenetic and synteny analysis supported *Pf*GNA as belonging to a gene family with a single origin early during the evolutionary diversification of the phylum. This Apicomplexa-specific GNA family grouped 6 additional orthologous sequences, including *Cryptosporidium parvum* and *Toxoplasma gondii*, which were also shown to complement the growth of the yeast temperature sensitive GNA mutant. Heterologous expression and *in vitro* activity assays confirm *Pf*GNA and *C. parvum* (*Cp*GNA) GNA enzymatic activities. The significance of GNA for the biochemistry of these parasites, together with the independent evolution and unique sequence features of this gene family in Apicomplexa, reveal GNA as a potentially relevant drug target amenable for selective inhibition.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-83

Conditional expression dilutes selection on developmental genes in the social amoeba Dictyostelium discoideum

Atahualpa Castillo Morales [1,*], Janaina Lima de Oliviera [1], Araxi Urrutia [1], Christopher Thompson [2], Jason Wolf [1]

[1]Department of Biology & Biochemistry and Milner Centre for Evolution, University of Bath, Bath, [2]Faculty of Life Sciences, University of Manchester, Manchester, United Kingdom

Abstract: Natural selection and genetic drift interact to shape patterns of molecular evolution, with effective population size typically being considered as the key factor limiting the role of selection. However, selection can generally only shape variation in coding regions when those regions are actually playing some functional role, and hence selection will be diluted for genes that are only expressed in certain contexts. This process will be particularly important for genes that are involved in phenotypes that are only expressed in some fraction of the generations, where we expect selection to be diluted by that fraction. For example, for most generations the social amoeba D. discoideum lives as a unicellular organism in soil undergoing vegetative growth, but when starved individuals aggregate into a multicellular structure and go through development to form a fruiting body. Given the conditional nature of this process, we expect developmental genes to display a signature of diluted selection compared to those that also play a role during vegetative growth. By defining developmental genes from expression profiles (high resolution developmental transcriptome) and estimating molecular evolution parameters derived from genome sequencing of natural isolates, we show in this study that conditionally expressed developmental genes appear to show similar patterns of molecular evolution as genes expressed during vegetative growth, but with a signature of relaxed selection. We have also tested whether conditional selection may limit the role of selection in shaping adaptive genomic features, such as patterns of GC content, and see similar signatures of diluted or relaxed selection. Our results support the hypothesis that conditional expression may dilute selection, shifting the balance of mutation and drift in shaping patterns of molecular variation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-69

Comprehensive Analysis of Evolutionary Constraint on the Spatial Distribution of Missense Variants in Human Protein Structures

R. Michael Sivley [1], William Bush [2], John A. Capra [3,*]

[1]Biomedical Informatics, Vanderbilt University, Nashville, [2]Institute for Computational Biology, Case Western, Cleveland, [3]Biological Sciences, Vanderbilt University, Nashville, United States

Abstract: The spatial distribution of genetic variation within protein structures is shaped by evolutionary and functional constraints and thus provides information about the importance of protein regions and the potential pathogenicity of protein alterations. We developed a statistical framework adapted from Ripley's K – a measure of spatial homogeneity used in spatial descriptive statistics – in which to model and compare spatial patterns of genetic variation across multiple distance scales. We then performed a comprehensive 3D analysis of protein-coding single-nucleotide variants (SNVs), including population-derived missense (ExAC, N=196,176) and synonymous (ExAC, N=122,963), pathogenic (ClinVar, N=4,827), and cancer (COSMIC, N=114,574) variants in 4,575 human proteins with solved structures in the Protein Data Bank.

Population-derived missense and synonymous, pathogenic, and cancer variants have drastically different spatial distributions. Missense variants observed in the general population are found in nearly every protein in our dataset (N=4,523) and are typically non-randomly distributed in the structure, while synonymous variants trend towards spatial randomness. In contrast, nearly 20% (88 of 453) of proteins with at least three pathogenic variants exhibit significant spatial clustering in 3D. This is substantially greater than the 0.3% (15 of 4,547) of proteins with significant clustering of cancer variants, in which there has been considerable recent interest. Indeed, we find that pathogenic variation is more clustered than recurrent cancer variation in two thirds (200 of 301) of proteins with sufficient occurrence of each.

Motivated by the significant 3D spatial clustering of pathogenic variation and the general dispersion of putatively benign variation, we hypothesized that analysis of variants' spatial distributions could provide a new line of evidence to use in the prioritization of variants of unknown significance (VUS). We will present several examples that demonstrate the strong predictive performance of our approach and its ability to complement common methods for the prioritization of candidate mutations. These findings suggest that explicit modeling of the spatial distribution of protein-coding variation has great promise to complement existing strategies for VUS prioritization and pathogenicity prediction.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-84

Structural drivers of protein family evolution in different eukaryotic lineages

Joseph B. Ahrens [1 2,*], Luis R. Nassar, Janelle Nunez-Castilla, Jessica Siltberg-Liberles

[1]Biomolecular Sciences institute, [2]Biological Sciences, Florida International University, Miami, United States

Abstract: Protein sequences often evolve with rate-heterogeneity among sites, where amino acid replacements occur much more frequently in some sites than in others. Previously, we conducted a large-scale study of metazoan protein families under an unbalanced factorial model to evaluate the relationship between site-specific evolutionary rates and three important factors: intrinsic disorder, secondary structure and functional domain involvement. Among other findings, our results indicated that, intrinsically disordered sites tend to evolve more quickly than ordered sites in agreement with earlier studies, but importantly, the overlap between the two rate distributions is large and there is a significant confounding interaction between disorder and secondary structure: disordered sites with high secondary structure propensity actually evolve slower on average than ordered, structured sites. To determine whether these trends are due to specific selective pressures in metazoans, or if they represent general trends in eukaryotic protein evolution, we have employed sequence-based computational methods to predict site-specific evolutionary rates and structural properties of over one million additional protein sequences from 15,404 plant, 6,358 fungal and 7,906 alveolate protist protein families. Here, we present the methodology and results of our effort to illuminate structural drivers of protein sequence evolution in eukaryotic lineages.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-80

Signatures of Positive Selection in Venom Proteins of Hawaiian Tetragnatha (Araneae :Tetragnathidae) Spiders

Timothy Cole [1,*], Michael Brewer [1]

[1]Biology, East Carolina University, GREENVILLE, United States

Abstract: The history of spider venom evolution has been marked by several events of gene duplications followed by neofunctionalization, allowing for the proliferation of a vast library of venom peptides across several spider taxa. Understanding the evolutionary pressures driving the diversification of venoms in spiders is difficult due to their ancient divergence times. Hawaiian *Tetragnatha* spiders present an ideal model for investigating the molecular evolution of spider venom proteins due the relatively recent adaptive divergences they have experienced with drastic niche shifts creating strong selective pressures on the venom. In this study we use transcriptomic data from venom glands of Hawaiian *Tetragnatha* spiders to present a method for characterizing the selective pressures of large biodiversity transcriptome datasets using a newly created bioinformatics pipeline. From eight individuals across five species 100,000 transcripts were grouped into 50,000 gene families by homology, 100 of those families were large enough to be subject to robust phylogenetic analysis and subsequent tests of pervasive positive selection. Two of those 100 families were found to be under strong positive selection at the amino acid level. The pipeline presented here was successfully detected signatures of positive selection in the venom sequences of a group of spiders that has undergone rapid adaptive radiations. Future work will expand the functionality of the proposed pipeline to manage sequencing data from a variety of biodiversity projects to better characterize broad patterns and processes of molecular evolution.

Expanded summary*: Recent advances in RNA seq have allowed for an abundance of coding sequences to be discovered at an incredible pace across all levels of biodiversity. The need for better bioinformatics tools to process this rapid influx of data has been the inspiration for the development of the pipeline we present, FUSTr (Families Under Selection in Transcriptomes). FUSTr was used to group transcripts, build gene family-level phylogenies, and test for pervasive positive selection amongst venom gene families in Hawaiin *Tetragnatha* spiders. Although, so far it has only been applied to spider venom transcriptomes the aim is to provide a tool that operates on any given coding sequence database.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-72

Clustering of amino acid substitutions in Drosophila, Primates, and Arabidopsis implicates positive selection and epistasis in evolution

Andrew Taverner [1,*], Peter Andolfatto [2]

[1]Quantitative and Computational Biology, [2]Lewis Sigler Institute, Princeton University, Princeton, United States

Abstract: Amino acid substitutions (AAS) are spatially clustered along the length of proteins for reasons that are as yet poorly understood. Variation in mutation rates and in selective constraints are obvious candidates. However, a large proportion of AAS between Drosophila (and other) species is inferred to be driven by positive selection. Recurrent adaptive substitutions, genetic hitchhiking, and epistasis may also contribute to clustering. A recent study by Callahan et al. used Drosophila species genome sequences to show that: 1) Clustering of AAS occurs on a length-scale of ~20 amino acid residues. 2) Clustering of AAS is stronger within than between lineages. 3) Charge-compensation is a significant a component of intra-lineage AAS clustering. Many aspects of biology differ among different groups of organisms. Here, we address whether the patterns from Drosophila generalize to two diverse taxa: primates and Arabidopsis.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP6

The interaction between selective constraints acting on substitutions and small-scale insertions and deletions in proteins

Marcin Bogusz [1,*], Simon Whelan [1]

[1]Ecology and genetics, Uppsala Universitet, Uppsala, Sweden

Abstract: Patterns of selection acting on genes can provide important insights into mechanisms of protein evolution. The most widely used method for studying selection is the ratio of synonymous to non-synonymous substitutions (dN/dS), which can distinguish neutral, purifying, and positive selection. This ratio, however, ignores the evolutionary forces acting on the other processes affecting protein evolution, including small-scale insertions and deletions (indels). Existing approaches for studying indel rates are typically based on filtered sequence alignments, and their estimates may be biased by gap patterns produced by alignment and filtering heuristics. As a consequence there is limited understanding of the how selection acts on indels and how it shapes protein evolution. We present a study that explicitly examines the relationship between selective constraint acting on substitutions (dN/dS) and the indel rate based on a new statistical method that simultaneously addresses both the alignment and phylogenetic problem. Our method jointly infers dN/dS and the indel rate from pairs of protein coding sequences, integrating across all possible pairwise alignments using an evolutionary model. We have examined ten thousand genes from human-mouse and human-chicken pairwise analyses, revealing that the indel rate and selection (dN/dS) tend to be correlated, demonstrating that purifying selection acting in proteins tends to affect non-synonymous mutations and indels in a similar way. The variation in indel rate, however, suggests there may be more complex evolutionary forces acting indels, although it cannot be accounted for by simple structural considerations, such as the number of domains. We have also investigated how the selective forces acting on substitutions and indels vary along genes. Our findings and methods offer the opportunity to begin studying the interaction between substitutions and indels, and the first steps towards understanding how they impact protein evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-104

Harnessing the genetic diversity of captive owl monkeys to study the special HIV variants that seed new human infections

Emily Feldman [1,*], Maryska Kaczmarek [1 2], Nicholas Meyerson [1], Cody Warren [1], Sandie Shan [1], John Nahabedian [3], Amit Sharma [3], Julie Overbaugh [3], Greg Wilkerson [4], Sara Sawyer [1]

[1]MCDB, BioFrontiers Institute, Boulder, [2]Ecology and Evolution, University of Texas at Austin, Austin, [3]Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, [4]Department of Veterinary Sciences, Michale E. Keeling Center for Comparative Medicine and Research, University of Texas MD Anderson Cancer Center, Bastrop, United States

Abstract: HIV-1 transmission to a new individual elicits a strong selective bottleneck on the existing virus population, and typically only one or a few very special viral variants establish each new infection. Only very recently have these transmitted/founder (T/F) HIV-1 variants been appreciated for their unique ability to initiate an infection. The vast majority of HIV-1 research has been performed utilizing lab-adapted or late-stage HIV-1 variants, which do not recapitulate the cell-tropism, unique resistance to interferon and restriction factors, and antigenic properties of T/F HIV-1 variants. Critically, the current animal model for HIV-1 infection and vaccine research, the rhesus macaque, does not support the replication of T/F HIV-1. Many of the T/F HIV-1 variants are highly restricted by the macaque version of CD4, the HIV-1 entry receptor. Additionally, macaques block T/F HIV-1 replication at stages post-entry due to several restriction factors with activity against HIV-1. Excitingly, our lab discovered the first monkey, the Spix's owl monkey, which encodes a CD4 that is broadly permissive for major global subtypes of T/F HIV-1. Further, owl monkeys are only known to encode a minimal number of restriction factors that block HIV-1, one of which (TRIM-Cyp) is evaded with a single point mutation in the HIV-1 genome, while another (tetherin) is polymorphic in owl monkeys with some individuals encoding ineffective alleles. I am harnessing the genetic diversity of multiple owl monkey colonies to define key molecular interactions of T/F HIV-1 biology. My preliminary data demonstrates that these studies will expand our understanding of T/F HIV-1 biology.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-99

Probing the relative accuracy of mutation--selection inference platforms

Stephanie Spielman*, Claus Wilke

Abstract: Achieving a thorough understanding of protein evolution relies on robust models that incorporate explicit mechanistic information. In recent years, the "mutation—selection" model of coding-sequence evolution has shown promise in reaching this goal. Mutation—selection models rest on fundamental population-genetics principles, and they estimate site-specific amino acid propensities and selection coefficient distributions. Two computationally-tractable inference approaches have been introduced: One employs a fixed-effects, highly-parameterized maximum likelihood framework (Tamuri et al. 2014 *Genetics*), while the other employs a random-effects Bayesian Dirichlet Process framework (Rodrigue and Lartillot 2014 *Bioinformatics*). While both implementations follow the same model, they make distinct predictions about the distribution of selection coefficients: The fixed-effects framework estimates a large proportion of highly deleterious substitutions, whereas the random-effects framework estimates all substitutions as either nearly-neutral or weakly deleterious. To obtain reliable inferences and justify a basis for future model development, we investigated the relative merits of these two approaches through a simulation-based strategy. We find that the fixed-effects approach, despite its extensive parameterization, consistently and accurately estimates site-specific evolutionary constraint. By contrast, the random-effects Bayesian approach systematically underestimates the strength of natural selection, particularly for slowly-evolving sites. We also find that, despite the salient differences between their inferences, the two approaches examined here yield surprisingly similar inferences of site-specific selective constraint. We conclude that the fixed-effects mutation--selection framework provides the more reliable software platform for model application and future development.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-93

Phylogenetic approaches for quantifying interlocus gene conversion

Xiang Ji [1 2,*], Alexander Griffing [1], Jeffrey Thorne [1 2]

[1]Bioinformatics Research Center, [2]Department of Statistics, North Carolina State University, Raleigh, United States

Abstract: Interlocus gene conversion (IGC) homogenizes repeats. While genomes can be repeat-rich, the evolutionary importance of IGC is poorly understood, largely due to a lack of statistical tools. We developed a strategy for adding IGC to widely-used probabilistic models for sequence change. In 14 groups of yeast ribosomal protein genes, we estimated the percentage of codon substitutions that originate with IGC rather than point mutation. We found values ranging from 20% to 38%. Our results are consistent with an important role for IGC in the evolution of each of the 14 protein coding gene families.

In this presentation, we summarize our approach and how we have been improving it by incorporating the distribution of IGC tract lengths and by increasing the number of paralogs per genome that can simultaneously be analyzed. We illustrate the enhancements with analyses of diverse data sets and we describe our efforts to characterize the enhancements via simulation. We also overview our study of whether some paralogs tend to be IGC donors rather than recipients, our work on how paralog divergence affects IGC rates, and our future plans. Taken together, our findings suggest IGC should and can be considered when multigene family evolution is investigated. This is especially relevant because IGC can affect paralogs created via whole genome duplication and so many key evolutionary lineages have experienced such duplications.

Expanded summary*: Interlocus gene conversion (IGC) homogenizes repeats. While genomes can be repeat-rich, the evolutionary importance of IGC is poorly understood, largely because of a lack of statistical tools. My first IGC publication (Ji et al. 2016. PMID 27297467) introduced an approach for characterizing it. The key idea was to jointly treat corresponding positions in different paralogs so that codon substitutions originating with both point mutation and IGC could be considered. We evaluated the approach with 14 data sets of yeast ribosomal protein genes and found the percentage of codon substitutions that originate with IGC rather than point mutation to range from 20% to 38%.

One might wonder whether this dramatic role of IGC is peculiar to certain yeast genes. Preliminary results from additional data are that IGC is abundant. For instance, our analysis of EDN/ECP data that was the basis of a classic study (Zhang et al. 1998. PMID 9520431) yields an estimate of >10% of substitutions from IGC. Whether 10% or 20% or 38%, these percentages are too high to safely ignore IGC, especially in light of the whole genome duplications experienced by so many evolutionary lineages. For such lineages, phylogenetic or divergence time inferences may be flawed if they disregard IGC. Plus, ignoring IGC is potentially problematic when reconstructing ancestral sequences or detecting positive selection. My work has relevance to these applications.

A shortcoming of our original approach was its inability to infer the tract length distribution of (fixed) IGC events. I developed a method to address this and am beginning to apply it. Motivated by primate ADH1 data, I extended our method so that up to 6 paralogs per genome can be considered. Also, I am improving my software in other ways (e.g., permitting asymmetry among paralogs in tendency to serve as IGC donor, investigating how distance between tandem paralogs affects IGC, summarizing how paralog divergence changes IGC). These efforts are technical in nature, but are important because they will enable others to answer questions that have previously been inaccessible. Finally, our aforementioned paper introduces a way to apply Felsenstein's pruning algorithm (the engine of interspecific likelihood inference) so that state spaces of size ten thousand are computationally tractable. We hope this spurs development of models that are more biologically plausible and useful than the 4-state and 61-state substitution models that predominate now.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-94

Evolutionary analysis of long disordered protein regions in humans

Toni Gossmann [1,*], Arina Afanasyeva [1 2 3], Chris Cooney [1]

[1]Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom, [2] Peter the Great St.Petersburg Polytechnic University, [3]Petersburg Nuclear Physics Institute (PNPI), St. Petersburg, Russian Federation

Abstract: Proteins are composed of functional units such as binding sites or catalytic domains and these are evolutionary well characterised. Much less is known about the protein parts outside of these functional units. A prominent example are disordered protein regions, e.g. protein parts that lack stable secondary or tertiary structures. Despite these unusual conformational features the vast majority of disordered protein regions are supposed to exert a function. Here we investigate this hypothesis by disentangling the selective forces that have acted within disordered protein regions. We systematically analyse more than 6000 human proteins with long disordered regions using a comparative phylogenetic approach with data from more than 90 mammalian species. We show that long disordered regions evolve more rapidly than their ordered counterparts. However, they are largely subject to purifying selection and are more likely to have experienced repeated episodes of positive selection. Within our approach we are able to pinpoint amino acid residues that potentially exhibit a functional role and are able to associate them with disease related variation in human populations. Taken together our results suggest that disordered protein regions are important targets of genetic innovation and have been largely overlooked from an evolutionary perspective in the past.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-102

Causes of single position fitness landscape changes

Anastasia Stolyarova [1,*], Elena Nabieva [2], Vasiliy Ptushenko [2], Georgii Bazykin [3 4]

[1]Faculty of Bioengineering and Bioinformatics, [2]Moscow State University, [3]Institute for Information Transmission Problems, Moscow, [4]Skolkovo Institute of Science and Technology, Skolkovo, Russian Federation

Abstract: Amino acid propensities at an amino acid site change with time, leading to amino acid substitutions and to molecular evolution. However, the causes of these changes are unclear. They may arise due to environmental changes; alternatively, even in a constant environment, they may arise due to changes in interacting sites elsewhere in the genome under epistasis. To distinguish between these possibilities, we analyze the phylogenetic distribution of substitutions at amino acid sites in the course of evolution of vertebrates, insects and fungi. We show that in general the fitness of the amino acid currently occupying a given amino acid site increases with time. Using simulations, we show that this is consistent with epistatic changes. By contrast, in rapidly evolving sites, the fitness of the amino acid currently occupying a given amino acid site declines with time, consistently with environmental fluctuations uncorrelated with the current genome content.

Statement: I would like to state my motivation to attend SMBE 2017 conference in Austin. I am a 6th year student of Faculty of Bioengineering and Bioinformatics at Lomonosov Moscow State University, where I study different aspects of modern biology. My areas of interest are evolutionary genomics and molecular evolution. I find them very interesting and challenging because in this field you can learn the basic principles of existence and development of all living organisms. The subject of my graduate thesis is also from this area and is devoted to dynamics of protein fitness landscapes. My long-term goal is to continue research in this area as PhD student. The conference theme is especially relevant to my professional expertise. I see this conference as the exceptional opportunity to network and build relationships with other scientists from all over the world and to see the people who've written important evolutionary papers I have read. The conference allows to see current hot topics in evolution research and to listen to new interesting studies. I also hope SMBE 2017 will give me the opportunity to receive experience in presenting my work and to get reliable response. Because the work language of the conference is English, I could naturally improve my English skills in practice. Also I would be happy to visit USA for the first time and see its nature and cities with my own eyes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP5

Progressively more subtle aggregation avoidance strategies form a long-term arrow of protein evolutionary time

Joanna Masel [1,*], Scott Foy [1 2], Ben Wilson [1 3], Rafik Neme [4], Matt Cordes [5]

[1]Ecology & Evolutionary Biology, University of Arizona, Tucson, [2]St. Jude Children's Research Hospital, Memphis, [3]AncestryDNA, San Francisco, [4]Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, [5]Chemistry & Biochemistry, University of Arizona, Tucson, United States

Abstract: Different gene families were born at different times, allowing us to compare young protein-coding genes to those that are older and "more evolved". To be retained during evolution, a protein must not only have a function, but must also avoid toxic misfolding. These two requirements are in conflict; hydrophobic amino acids form the cores of protein folds, but also promote aggregation.

Young genes have a hydrophilic amino acid composition, and statistically controlling for protein length and evolutionary rate shows this is not the result of homology detection bias. Young gene hydrophilicity exceeds that of intergenic controls as well as that of old genes, contradicting a recent "continuum" theory of de novo gene birth and confirming an alternative theory of "preadaptation".

Hydrophobic amino acids were previously believed to be excessively interspersed along the primary sequence, compared to scrambled sequences. We found that this is only true for the very oldest genes with homologs in prokaryotes; younger genes have progressively higher levels of clustering of hydrophobic residues as a function of youth. This suggests a long-term arrow of protein evolutionary time driven by a gradual shift in strategy for avoiding misfolding over billions of years. The few hydrophobic residues within young genes are clustered near one another along the primary sequence, presumably to assist folding. Old genes avoid clustering, presumably to better avoid misfolding while finding more subtle ways to fold, and can tolerate higher total hydrophobicity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-91

Comparative population genomic analysis of the factors driving fitness effects of protein-changing mutations

Christian D. Huber [1,*], Bernard Kim [1], Clare D. Marsden [1], Kirk E. Lohmueller [1 2 3]

[1]Department of Ecology and Evolutionary Biology, [2]Interdepartmental Program in Bioinformatics, [3]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, United States

Abstract: Protein-changing mutations affect the function, folding and stability of proteins, and consequently, may affect individual fitness. However, the evolutionary mechanism determining the distribution of fitness effects (DFE) of mutations in natural populations remains unclear. Moreover, models from evolutionary biology, systems biology, and protein biophysics emphasize different factors such as protein function and stability, phenotypic complexity, and mutational robustness of gene regulatory networks in determining the DFE. Here, we develop a novel comparative population genomic framework that tests whether the DFE systematically differs across species. We show that when comparing yeast, *Drosophila*, mice and humans, the average fitness effect of a protein-changing mutation becomes more deleterious with increasing species complexity. Additionally, genes with tissue-specific expression patterns show more variable mutational effects on fitness than broadly expressed genes. Comparing five theoretical models for the evolution of the DFE suggests that Fisher's Geometrical Model fits best to the data, indicating that phenotypic complexity and population size are the predominant drivers of DFE differences between species. Our results suggest that gene expression level and tissue specificity are major determinants of differences in the DFE among genes. Furthermore, selection on protein-changing mutations acts differently across phylogenetically divergent species, affecting lineage specific divergence and thus contributing to the overdispersion of the molecular clock. We lay the foundation for future work that could include developing more complex models of selection, and improved prediction of disease-causing variants.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-90

The evolution of a multi-step enzyme in the corticosteroid pathway in primates

Carrie Olson-Manning [1,*], Joseph Thornton [2]

[1]Biology, Augustana University, Sioux Falls, [2]Ecology and Evolution, University of Chicago, Chicago, United States

Abstract: Gene duplication is a major driver of the augmentation and functional evolution of biochemical pathways. Here we study the evolution of an enzyme family that performs multiple enzymatic reactions in the corticosteroid pathway in primates to understand how duplication of its constituent enzymes evolved after a gene duplication. The corticosteroid pathway is phylogenetically diverse, with several gene duplications in mammals. The products of the pathway (aldosterone, corticosterone, and cortisol) are steroid hormones in tetrapods that regulate metabolism and stress. We study a duplication in the primate lineage of the multi-step enzyme, Cytochrome P450 11B (CYP11B), that synthesizes these hormones. We find that one copy in humans (CYP11B1) specializes on production of cortisol and its paralog (human CYP11B2) specializes on producing aldosterone. With ancestral state reconstruction and heterologous expression we find that ancestor of CYP11B1/2 had moderate ability to synthesize cortisol and aldosterone, consistent with the escape from adaptive conflict and functional specialization following a duplication. We dissect the functional effects of the substitutions responsible for the functional shift in the CYP11B2 lineage and find that all substitutions in this lineage affect function with a mix of additive and epistatic effects. These results suggest that gene duplication can lead to pathway elongation when the paralogs of a multi-step enzyme specialize on different pathway products, especially when there is a negative tradeoff between activity on each enzymatic step.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP7

Contingency and evolvability across the sequence space of an ancestral protein

Tyler Starr [1,*], Joe Thornton [2]

[1]Biochemistry and Molecular Biology, [2]Ecology & Evolution and Human Genetics, University of Chicago, Chicago, United States

Abstract: To understand how evolution produced the specific proteins of the present, we can reconstruct the history of their sequences and functions; to understand why it did so, we must compare the path that evolution took to alternative paths that could have been but were not taken. We used high-throughput mutational scanning to characterize alternative histories in the sequence space of an ancient transcription factor, which evolved a novel specificity for DNA binding sites through well-characterized mechanisms. We assayed all combinations of all amino acids at four key sites in ancestral proteins for their ability to bind ancestral and derived DNA binding sites. We found hundreds of genotypes that recapitulate the historical transition in function through diverse biophysical mechanisms. Alternative outcomes, like the historical one, require permissive substitutions to be accessed from the ancestral protein, but not all require the same permissive substitutions that occurred in history. Had evolution began from other genotypes with the ancestral function, the derived function would have still been accessible. However, alternate starting points would have reached outcomes with different underlying forms than the historical outcome, and some alternate starting points would not have required permissive substitutions. Our results illuminate the vast sequence space of possibilities from which evolution sampled one path, highlighting the stochasticity on which the outcomes of evolution depend.

Expanded summary*: The evolutionary process produces contingent histories in which just one or sometimes several outcomes are realized out of a potentially large realm of alternatives. As such, dissections of historical evolution impose blinders on our view of sequence-function landscapes and the mechanisms that characterize evolution across these landscapes. Broader insight into the sequence-function landscapes navigated by evolution can be achieved by comparing the paths taken in evolution to alternative paths that could have been but were not taken. Knowledge of these alternate histories would reveal whether the specific outcome of evolution represents the only or optimal way to evolve a particular function, or whether it was just one of many possibilities; whether or not permissive substitutions that are neutral with respect to the derived function had to be stochastically acquired to make the derived sequence or function "evolvable"; and whether or not the derived form and function could have evolved from other genotypes with the ancestral function. These issues of contingency, epistasis, and evolvability – and how they manifest in the sequence-function landscapes traversed in evolution – are of fundamental importance to our understanding of the molecular evolution of proteins and other biomolecules. They also impact various applications that depend on the structure of the sequence-function landscape, such as protein engineering, evolutionary forecasting, and predictions of the phenotypic or pathogenic effects of genetic variation.

To characterize the sequence-function landscape over which a historical evolutionary transition in function occurred, we applied high-throughput mutational scanning to ancestral steroid receptor transcription factors bracketing a historical transition in DNA-binding specificity. We experimentally characterized all amino acid combinations at four key residues at the protein-DNA interface through a high-throughput FACS-seq approach. Our experiments reveal many alternative outcomes that recapitulate the historical change in function, suggesting the historical outcome was arrived upon stochastically. The derived function is highly evolvable, as it is accessible from virtually anywhere in the neutral network of the ancestral function. Nonetheless, the types of trajectories that could be followed are highly contingent on starting point: for example, the historical trajectory was dependent on epistatic permissive substitutions and likely passed through a promiscuous intermediate, but alternative trajectories can proceed through discrete shifts that are driven solely by selection for the derived function. This analysis distinguishes mechanisms of functional evolution that are general across sequence space from those that are peculiar to the historical trajectory. More broadly, our results emphasize how the multidimensional nature of sequence space facilitates evolutionary transitions in function, resulting in evolutionary histories that may be largely driven by chance historical events.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-89

An unbiased view of functional shifts and co-evolution across large protein families: Integrating computational and experimental approaches

Bryan Kolaczkowski [1,*], Charles Pugh [1], Kelsey Aadland [1], Raquel Dias [1]

[1]Microbiology & Cell Science, University of Florida, Gainesville, United States

Abstract: Determining the evolutionary forces and contexts driving transitions in protein function, the structural mechanisms through which proteins change their functions and the effects of functional changes on cellular signaling networks is central to molecular evolution. Although these questions have been widely and deeply examined using computational and statistical approaches, experimental characterization of protein functional evolution is typically confined to short-term, highly-controlled conditions which may not extrapolate to long-term natural evolutionary processes. Ancestral sequence resurrection (ASR) studies have combined statistical inference of ancient protein sequences with experimental characterization of these sequences in order to directly identify how evolutionary changes in protein sequence impact structure and function. However, the reliance of these studies on low-throughput experimental approaches has limited them to examining a relatively small number of hypotheses, which are typically determined based on phylogenetic patterns and/or limited characterization of protein function in model systems. Currently, we have very little direct, unbiased information about how protein function may change across large gene family trees. Here we develop an approach that integrates large-scale ancestral sequence reconstruction with structural modeling, molecular dynamics, statistical machine learning and experimental validation to characterize molecular-functional shifts and their structural bases across gene family trees with 1000s of sequences. We apply this approach to the study of two families of RNA-binding proteins spanning animal and plant lineages. We identify a number of discrete shifts in protein-ligand affinities across individual branches on the tree as well as long-term changes in ligand affinity occurring slowly across multiple nodes. Interestingly, our results suggest that changes in protein molecular function may not always be associated with gene duplication or major speciation events, suggesting that proteins may change their functions more often than we might suspect. The identification of functional 'flip flopping' - repeated transitions among a small number of functional states through different structural mechanisms - also supports the view that protein function may be highly evolutionarily labile. We identify and characterize long-term co-evolution between RNA binding proteins and their downstream signaling partners, suggesting that co-evolutionary processes may also be common, even in cases where large shifts in molecular function are rare. We suggest that an 'unbiased' view of protein functional evolution may reveal new information about how protein families evolve when we aren't looking.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP9

Massively-parallel bacteriophage evolution quantifies the adaptive utility of genetic codes expanded with nonstandard amino acids

Colin Brown [1 2,*], Michael Hammerling [3], Jeffrey Barrick [1 2]

[1]Institute for Molecular and Cell Biology, [2]Molecular Biosciences, University of Texas at Austin, Austin, TX, [3]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, United States

Abstract: The existence of natural genetic codes with 21st amino acids suggests that evolution can benefit from encoding side-chain chemistries not present in the standard alphabet. The evolutionary processes that drive genetic code expansion, however, remain poorly understood. The recent development of amber-suppression systems that efficiently incorporate nonstandard amino acids (nsAAs) into proteins, along with genomically recoded organisms with no native amber stop codons, now allows the evolutionary potential of non-standard alphabets to be tested directly. We used whole-genome sequencing of a highly parallel experiment with bacteriophage T7 to examine how evolution of its proteome differed in *E. coli* hosts that recoded the amber codon to one of five nsAAs or to the canonical amino acid tyrosine. In total, hundreds of thousands of independent mutations were observed, including nearly a thousand mutations to nsAAs. We found that phage populations propagated on genetic codes with certain 21st amino acids evolved to utilize them at the same rate as the canonical amino acid control, but with unique substitution spectra. Mutations to nsAAs were found multiple times at specific sites in bacteriophage genes critical for infection, replication, and capsid packaging. This genetic parallelism suggests that these substitutions were beneficial to phage fitness. We further characterized one such mutation in T7 RNA polymerase. Computational structural modeling and biochemical assays demonstrated that substitutions of nsAAs at this site had effects that were not possible with a standard amino acid alphabet. Our results highlight the potential for genetic codes expanded with unique chemical functionalities to improve adaptation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP1

Quantitative models of protein evolution informed by experimental data

Jesse Bloom*

Abstract: Over the last few decades, the study of protein evolution has proceeded along two largely independent tracks: one molecular, and the other statistical. From a molecular perspective, experiments have made detailed case studies of the constraints that shape protein evolution. For instance, experiments have illuminated a great deal about how selection for structure and other biophysical properties shape protein evolution, and have yielded at least some understanding of the nature of epistasis among intra-protein mutations. From a statistical perspective, mathematical models of sequence evolution have been formulated for the quantitative analyses of sequences in phylogenetics and population genetics. These models have become increasingly statistically sophisticated, but make relatively little use of the detailed molecular knowledge that has been gleaned from experiments.

Recently new high-throughput experiments have begun to provide sufficiently comprehensive information that it is now possible to inform statistical models of evolution with protein-specific molecular information. I will discuss such models, which merge molecular and statistical perspectives on protein evolution. I will focus on the following questions:
1) What is the right way to incorporate experimental data into quantatitive models of protein evolution?
2) Experiments are typically done on a single specific protein, but evolutionary models seek to span a set of homologs -- how can we reconcile this gap?
3) How can we use experimental data to aid in common statistical goals such as detecting adaptive evolution from sequence data?
4) How can we statistically quantify the relative accuracy or absolute adequacy with which experiments capture the actual constraints on a protein in nature?

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP2

Evolution of novel proteins by gene remodeling in animals.

Mary O'Connell [1,*]

[1]Biology, University of Leeds, Leeds, United Kingdom

Abstract: To follow.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-62

Statistically assessing the relevance of experimentally defined immune epitopes to natural viral evolution

Sarah Hilton [1,*], Jesse Bloom [2 3]

[1]Genome Sciences, University of Washington, [2]Basic Sciences, [3]Computational Biology, Fred Hutchinson Cancer Research Center, Seattle, United States

Abstract: Defining the regions of a viral protein targeted by the immune system is important for understanding virus evolution. While there are many experimental methods to define such immune epitopes, it is difficult to evaluate how well these definitions represent the targets of anti-viral immune selection in nature. To objectively assess the evolutionary relevance of experimentally defined epitopes, we have developed a phylogenetic framework to quantify how well different epitope definitions describe the natural immune selection on a virus. Specifically, this framework allows pre-specified epitope sites to evolve under a different rate of nonsynonymous-to-synonymous mutations than other sites. Since epitopes are often under positive selection, we expect models that correctly identify these epitopes to outperform a null model that assigns one nonsynonymous-to-synonymous ratio to all sites. We used our approach to assess five definitions of immune epitopes for influenza virus's hemagglutinin protein. Three of the five epitope sets significantly outperformed the null model when applied to human influenza. In contrast, the epitope definitions generally led to less improvement when applied to swine influenza, which is thought to be under less immune selection. As experimental methods to map viral epitopes become more comprehensive and quantitative, our statistical approach will provide a framework for evaluating how well these experiments describe selection on viruses in nature.

Expanded summary*:

Defining the regions of a viral protein targeted by the immune system is important for understanding virus evolution. While there are many experimental methods for defining such immune epitopes, it is difficult to evaluate how well these experimental definitions represent the targets of anti-viral immune selection in nature.

To objectively assess the evolutionary relevance of experimentally defined epitopes, we have developed a phylogenetic framework to quantify how well different epitope definitions describe the immune selection on a virus in nature. We define a null model based on the Halpern-Bruno mutation-selection model with the site-specific amino-acid preferences from deep mutational scanning experiments and a gene-wide rate of nonsynonymous-to-synonymous mutations. Since epitopes are often under positive selection and have a higher rate of nonsynonymous-to-synonymous mutation than other sites in a protein, we define an alternative model with site-specific rates of nonsynonymous-to-synonymous mutations dependent on the epitope definition. Now, epitope definitions represent the *a priori* expectation of strength with which a given site is thought to be targeted by immunity. We can then statistically compare the null model and the alternative model to test how well the epitope definition describes immune pressures in nature. We expect only the alternative model with correctly identified epitopes to significantly outperform the null model.

We used our approach to assess five definitions of immune epitopes for influenza A virus's hemagglutinin protein. For each definition, every site in the protein is classified as either an epitope or a non-epitope site. Three of the five epitope definitions significantly outperformed the null model when applied to human influenza. In contrast, the epitope sites generally led to less improvement when applied to swine influenza, which is thought to be under less immune selection than human influenza. This method can be used to evaluate both binary epitope definitions, such as these five examples, and continuous epitope definitions. This flexibility is important as experimental methods to map viral epitopes become more comprehensive and quantitative. Overall, our statistical approach provides a framework to objectively evaluate how well experimentally defined epitopes describe selection on viruses in nature even as the complexity of epitope definitions increases.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP8

Quantifying the divergence in mutational effects among distant homologs of HIV's envelope protein

Hugh Haddox [1 2,*], Adam Dingens [1 2], Jesse Bloom [2]

[1]MCB PhD Program, University of Washington, [2]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, United States

Abstract: As homologous proteins diverge in sequence, the mutations that they can tolerate also change. Such shifts can arise from epistatic interactions among residues, where the amino-acid identity at one residue affects which mutations are tolerated at the other. Such epistasis can dramatically affect adaptive evolution. However, since studies of epistasis typically focus on small subsets of sites, we lack a comprehensive understanding of its prevalence protein-wide. In particular, it is unclear both how many sites change in mutational tolerance as homologs diverge and the typical magnitude of these changes. Using deep mutational scanning, we have experimentally estimated the effects of all single amino-acid changes to two homologs of HIV's envelope protein that are 77% identical. For each homolog, this involved making a library of genes with all single amino-acid mutations, selecting for variants that could support viral replication in cell culture, and deep sequencing the library before and after selection to quantify the degree each amino acid was tolerated at each site. We found that a substantial number of sites (40%) significantly differ in mutational tolerance between homologs, with most differences corresponding to small-to-intermediate effect sizes. We are currently extending this experiment to a third homolog to test the generalizability of our findings. Overall, this work helps elucidate both the frequency and strength of epistasis between homologs protein-wide.

Expanded summary*: As homologous proteins diverge in sequence, the mutations that they can tolerate also change. Such shifts can arise from epistatic interactions among residues, where the amino-acid identity at one residue affects which mutations are tolerated at the other. Numerous studies have found that epistasis can dramatically affect protein evolution by opening or closing the door to adaptive mutations. However, since these studies typically focus on small subsets of sites, we still lack a comprehensive understanding of its prevalence protein-wide. In particular, it is unclear both how many sites change in mutational tolerance as homologs diverge and the typical magnitude of these changes.

Using a high-throughput technique called deep mutational scanning, we have experimentally estimated the effects of all single amino-acid changes to two homologs of HIV's envelope protein (Env) that are 77% identical. To do so, for each homolog, we first made a library of *env* genes with all single amino-acid mutations. Next, we selected for variants that could support viral replication in cell culture. Since Env's receptor-binding and membrane-fusion activities are essential for viral replication, we expected this selection step to enrich functional variants and deplete nonfunctional variants. Finally, we deep sequenced the library before and after selection to quantify the enrichment or depletion of each mutation. The results provide a quantitative estimate of how well each of the 20 amino acids is tolerated at each site in both homologs. Using experimental replicates to account for noise, we found that a substantial number of sites (40%) significantly differ in mutational tolerance between homologs, with most differences corresponding to small-to-intermediate effect sizes.

Overall, this work helps elucidate both the frequency and strength of epistasis between homologs protein-wide. This approach has also been used to compare the mutational tolerance of two homologs of influenza virus's nucleoprotein that are 96% identical. For these closely related homologs, only a small fraction (6%) of sites were found to significantly differ in mutational tolerance between homologs, with correspondingly small shifts in mutational tolerance. Our study of Env expands upon this work by providing an example of more divergent homologs where epistasis is both more frequent and of greater strength. To assess whether our findings change at different levels of sequence-level divergence, we are currently extending this experiment to a third Env homolog that is 77% and 88% identical to the first two.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP13

Does adaptive evolution proceed by small or large steps at the molecular level?

Juraj Bergman*, Adam Eyre-Walker [1]

[1]School of Life Sciences, University of Sussex, Brighton, United Kingdom

Abstract: Whether adaptation proceeds by small or large steps is a fundamental problem in evolutionary biology. We have investigated this question at the molecular level by estimating the rate of adaptive evolution for every pair of amino acids separated by a single mutational step. We conduct our analysis by using an extension of the McDonald-Kreitman framework, divergence between *Drosophila melanogaster* and *D. yakuba*, and polymorphism data from 200 African *D. melanogaster* lines. We find that the rate of adaptive evolution is higher between amino acids that are more similar for a broad range of physiochemical properties, suggesting that small changes are more common in adaptive evolution. However, the slope of the relationship between the rate of adaptive evolution and the difference between amino acids is such that the total amount of adaptive evolution contributed by large changes is greater than that contributed by small changes. Adaptive evolution at the molecular level therefore seems to proceed by larger, not smaller, steps.

Expanded summary*: In many species, there is now good evidence of genome-wide adaptive evolution in protein-coding sequences. Amino acid substitutions affect protein properties in various ways and at different scales. This is due to the fact that any two amino acids differ (in various degrees) with respect to their mass, polarity, charge, volume and other attributes. However, a substitution which occurs between amino acids that are similar in their physiochemical characteristics will likely not affect protein function in a major way, and can therefore be considered conservative. Conversely, some amino acid substitutions are likely to be radical with respect to certain physiochemical properties. Furthermore, the likelihood of conservative and radical amino acid substitutions is dependent upon their rate of occurrence and the potential to contribute to adaptive change. The question which naturally arises from these considerations is: does adaptation generally proceed by conservative or radical amino acid substitutions - *i.e.*, is adaptive evolution taking large or small steps?

To address this question of significant biological importance, we investigate the relationship between step size and adaptive evolution by analyzing segregating amino acid pairs underlied by a single mutational step, observed in a population dataset of African *Drosophila melanogaster*, consisting of almost 200 whole genome sequences. For each amino acid pair, we infer the corresponding site frequency spectrum (SFS) using *D. melanogaster* polymorphism data, as well as nucleotide divergence using *D. yakuba* as an outgroup. By applying a variant of the McDonald-Kreitman analysis, which takes into account the effect of slightly deleterious mutations on the SFS, we infer the rates of adaptive evolution and the proportions of substitutions fixed by positive selection for each observed amino acid pair. These estimates, together with the abundance of available data on different amino acid similarity indices, form the basis of our analysis and allow us to quantify the step size of adaptive evolution.

Several important conclusions follow from our analysis. Firstly, while the rate of adaptive evolution is faster between more similar amino acids, the proportion of substitutions fixed by positive selection is equal for both radical and conservative amino acid pairs, indicating that adaptive radical mutations confer greater fitness advantages. Furthermore, the analysis of correlations between the rate of adaptive evolution and various amino acid similarity matrices reveals that conservative mutations occur more frequently in adaptive evolution. Consistent with this notion, we observe a strong positive correlation between the rate of adaptive evolution and amino acid substitution rates as defined by the BLOSUM matrices. We also recover the strong negative correlations between the rate of adaptive evolution and the two canonical amino acid distance matrices as defined by Grantham (1974) and Miyata (1979). Finally, to quantify the step size of adaptive evolution, we conduct a linear regression analysis and observe that the increase of amino acid distance is, on average, associated with a correspondingly lower change in the rate of adaptive evolution. Taken together, our results clearly indicate that, despite the higher rates of conservative evolution, the majority of adaptive change at the molecular level is contributed by mutations with larger, rather than smaller, effect sizes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-65

Structural insights into abalone egg-sperm interactions and the possible limits of X-ray crystallography in studying rapid evolution

Damien Wilburn [1,*], Willie Swanson [1]

[1]Genome Sciences, University of Washington, Seattle, United States

Abstract: A hallmark of reproductive proteins that mediate egg-sperm interactions is rapid evolution. Strong selection to maintain successful fertilization coupled to differences in male/female reproductive strategies often promotes arms race dynamics that drive accelerated evolution to maintain high affinity protein-protein interactions. While reproductive isolation often occurs through changes in timing or location of reproduction, the continual co-evolution and molecular refinement of gamete recognition proteins can create boundaries to hybridization. For the marine gastropod abalone, seven sympatric species with similar breeding seasons live off the coast of California, yet hybrids are rarely observed. During fertilization, abalone sperm secrete a 16 kDa acrosomal protein, lysin, which specifically binds to repeat domains in the egg coat protein VERL. Lysin-VERL interactions are species specific, and molecular evolutionary studies demonstrate strong signatures of positive selection and rapid co-evolution between the two proteins. However, lysin acquires approximately five times as many non-synonymous substitutions as VERL, and the molecular mechanism of how these mutational effects contribute to high-affinity, species specific interactions remains unclear. Using multidimensional NMR, we solved the solution structure of lysin from red abalone (*Haliotis rufescens*). Multiple lines of evidence support the coordination of the N- and C-terminus, forming a nexus of positively selected sites that constitute a likely VERL-binding interface. Biochemical experiments coupled with docking simulations suggest that lysin transitioned from a monomer to dimer in response to increased duplication of VERL repeats. Notably, sites under positive selection were found to be enriched on regions with increased backbone flexibility, and these regions of positively selected sites were heavily distorted in the lysin crystal structures. Comparison of additional rapidly evolving systems suggest that such crystal artifacts may be widespread, highlighting the importance of rapidly evolving proteins likely having unique physiochemical constraints and structural features.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

OM-MP14

Structure of global epistasis in protein sequence-function relationships

Jakub Otwinowski [1,*], David McCandlish [2], Joshua Plotkin [1]

[1]Biology, University of Pennsylvania, Philadelphia, [2]Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, United States

Abstract: The evolution of proteins depends on their fitness landscapes, which may be extremely complex due to genetic interactions. By fitting models to data from high-throughput sequence-function assays, we hope to elucidate the structure of these epistatic interactions and to make predictions of the protein's evolution. Unfortunately, standard approaches which specify pairwise epistasis have thousands of parameters and often extrapolate poorly. I show how deviations from a linear model suggest a new class of statistical models, in which sites are coupled via a non-linear function of the sum of additive effects, representing global epistasis. These models are comprehensible, in that they infer an intermediate phenotype, which may be predictive of underlying molecular phenotypes, such as protein stability or functional binding, and these fitness landscapes are single-peaked when the non-linear global coupling is monotonic. Our statistical framework quantifies the amount of uncertainty due to measurement noise and randomness due to non-global epistasis. It can also test different variants of the model to decide which one best describes the data with the fewest parameters. I show how our model fits data from protein mutagenesis studies better than previous models, with far fewer parameters, and while making better predictions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-96

Towards Understanding the Biophysical Foundation of Trimethoprim Resistance Evolution in Escherichia coli

Yusuf Talha Tamer [1,*], Erdal Toprak [1]

[1]Green Center for Systems Biology, University of Texas Southwestern Medical Center, Dallas, United States

Abstract: Antibiotic resistance is a growing public health problem. Acquisition of resistance via spontaneous mutations in the drug target genes is one of the common mechanisms of antibiotic resistance evolution. We have previously observed acquisition of up to five resistance-conferring mutations in dihydrofolate reductase (DHFR) when we evolved initially drug sensitive E. coli cells in the morbidostat using trimethoprim, an antifolate drug. DHFR is a ubiquitous enzyme that produces an essential metabolite for cell growth and maintenance by reducing 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate (THF) by hydride transfer from NADPH. Trimethoprim molecules competitively bind DHFR and block DHFR's enzymatic activity. To be able to understand how DHFR preserve its catalytic activity while accumulating multiple resistance-conferring mutations in the catalytic core of DHFR, we have purified and characterized all combinations of six resistance-conferring mutations. By using the biochemical parameters of these mutants, we have run computer simulations to identify most plausible genetic pathways that reach highest possible trimethoprim resistance without sacrificing catalytic power. Finally, we have quantified epistatic interactions between resistance-conferring mutations. By creating a deeper understanding for the evolutionary dynamics of an important drug target enzyme, our work provides experimental and computational tools for studying protein evolution with the ultimate goal of improving human health.

Expanded summary*: Antibiotic resistance evolution is a growing public health problem. In United States, each year more than two million of people are infected with multi-drug resistant bacteria and more than 23,000 of them are lethal and unfortunately these numbers are increasing[1]. Our battle with antibacterial resistance has started with our first paper that was published in Molecular Biology and Evolution journal. In this work, we showed the cross resistances between antibiotics. Our major finding was being resistant to aminoglycoside class of antibiotics makes bacteria hypersensitive to other classes of antibiotics such as folate synthesis pathway inhibitors or 50S ribosomal inhibitors [2]. Then we further pursued the phenomenon of hypersensitivity and checked the genomic changes in these hypersensitive bacteria. We found out that although specific point mutations in efflux pumps makes bacteria resistant to aminoglycosides, it also increases the sensitivity to many others. In following study, we combinatorically deleted a set of efflux pumps in *E. coli* and measured their resistances to 27 different antibiotics and found out that AcrAB-TolC pump system is one of the main players for resistance evolution. Any disruption in this pump system is increasing bacterial sensitivity to multiple antibiotic classes [3].

Previous studies we have done was targeting nonspecific resistance evolution. Then we wanted to understand a target specific resistance evolution in biophysical and biochemical levels since this case also have an aspect on protein evolution. Thus, we used evolution of trimethoprim resistance, a widely-used synthetic antibiotic that competitively inhibits dihydrofolate reductase (DHFR) as a model system. This model is suitable for us since evolution of trimethoprim resistance happens in multi-step fashion and inherited changes in bacteria are focused on the target enzyme as our whole genome sequencing results of trimethoprim resistant bacteria showed. Previous works have revealed the resistance conferring mutations but the effects of these mutations are not studied in biochemical level. We have purified and characterized all combinations of these resistance conferring mutations up to five mutations. Our results show that each mutation is increasing resistance with a cooperative manner but survival is also depending on preserving the catalytic activity. Interestingly, we observed a bifurcation in the catalytic activities of these mutants. Although each mutation has a subtle decreasing effect on catalytic activity by themselves, acquiring one specific mutation (P21L) on top of 2 or more mutations in the background is killing the catalytic activity up to 250-fold. We also calculated the epistasis between these mutations. Ensemble averaged epistasis values are increasing with the increasing number of mutations meaning the system has a cooperative network both for catalysis and resistance. To further understand this phenomenon and to find out the plausible trajectories that can be taken by evolution, we used our biochemical data as a measure of fitness and we simulated the evolution of DHFR. Understanding the structural level is also a part of this project. To this end we started a collaboration to study the effects of accumulating resistance-conferring mutations in structural level using MD simulations. Results that we obtained from these simulations reveal modifications in H-bonding network between ligand, and enzyme and changes in catalytically active loop motions.

1. *About Antimicrobial Resistance*. 2015 [cited 2017 02.01.2017]; Available from: https://www.cdc.gov/drugresistance/about.html.
2. Oz, T., et al., *Strength of selection pressure is an important parameter contributing to the complexity of antibiotic resistance evolution.* Mol Biol Evol, 2014. **31**(9): p. 2387-401.
3. Ayhan, D.H., et al., *Sequence-Specific Targeting of Bacterial Resistance Genes Increases Antibiotic Efficacy.* PLoS Biol, 2016. **14**(9): p. e1002552.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-97

Inference of Amino Acid functionality from DNA sequences using a novel phylogenetics approach

Cedric Landerer [1,*], Jeremy Beaulieu [2], Brian O'Meara [1], Michael Gilchrist [1]

[1]Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Knoxville, [2]Biological Sciences, University of Arkansas, Fayetteville, United States

Abstract: Understanding how the amino acid sequence impacts protein functionality is a fundamental question in molecular biology and evolution.
We present a novel approach that allows the estimation of amino acid functionality from sequence alignments
and incorporates sequence relatedness using phylogenetics.
Most phylogenetic approaches rely on time reversible models with symmetric substitution rate matrices.
In contrast, we employ a non-symmetric substitution rate matrix based on models of population genetics,
incorporating mutation, selection and genetic drift.
We assume selection is proportional to the distance between observed amino acids and the optimal amino acid in physicochemical space,
which allows for the estimation of a functionally optimal amino acid sequence given the aligned DNA sequence and their relatedness.
Other phylogenetic approaches that include information on amino acid functionality require expensive and time consuming deep mutation scanning (DMS) experiments.
However, DMS is limited to fast growing organisms where mutation libraries and targeted artificial selection can be applied.
DMS also oversimplifies the process of introduction and purging of variation by mutation, selection and genetic drift.
The presented approach can be applied without restriction to all protein coding genes where information on relatedness is available.
We demonstrate that our method out performs other phylogenetic models that incorporate DMS data using DMS data for E. coli beta-lactamase.
Furthermore, we present an extension allowing for changes in the optimal amino acid along the phylogenetic tree and demonstrating its effectiveness using influenza.
Understanding how the amino acid sequence impacts protein functionality is a fundamental question in molecular biology and evolution.We present a novel approach that allows the estimation of amino acid functionality from sequence alignmentsand incorporates sequence relatedness using phylogenetics.Most phylogenetic approaches rely on time reversible models with symmetric substitution rate matrices.In contrast, we employ a non-symmetric substitution rate matrix based on models of population genetics,incorporating mutation, selection and genetic drift. We assume selection is proportional to the distance between observed amino acids and the optimal amino acid in physicochemical space,which allows for the estimation of a functionally optimal amino acid sequence given the aligned DNA sequence and their relatedness.Other phylogenetic approaches that include information on amino acid functionality require expensive and time consuming deep mutation scanning (DMS) experiments.However, DMS is limited to fast growing organisms where mutation libraries and targeted artificial selection can be applied.DMS also oversimplifies the process of introduction and purging of variation by mutation, selection and genetic drift.The presented approach can be applied without restriction to all protein coding genes where information on relatedness is available.We demonstrate that our method out performs other phylogenetic models that incorporate DMS data using DMS data for E. coli beta-lactamase.Furthermore, we present an extension allowing for changes in the optimal amino acid along the phylogenetic tree and demonstrating its effectiveness using influenza.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mechanisms of protein evolution

POA-98

Evolution of regeneration in animals

Pinglin Cao [1,*], Takashi Makino [1]

[1]Graduate School of Life Sciences, Tohoku University, Sendai, Japan

Abstract: From microorganisms to mammals, all species have regenerative ability, which is a capacity for regrowing or restoring the lost body part. Some species such as flatworms, which can regenerate a new individual from a small body fragment, have strong regenerative ability, have strong regenerative ability, but other ones appear to have been greatly restricted. It has been reported that only a few genes (CAP-43, homeobox *msx*, Wnt, etc.) are related to regenerative ability (Fu et al., 1997; Liu et al., 2013; Akimenko et al., 1996), although many of genes may be involved in developing and resorting the lost body part during the regenerative process. Furthermore, the evolution of genes with regenerative ability among the animals is still unclear. We hypothesize that there are master genes related to regeneration in animals, and losses of them or nosynonymous substitutes few amino acids, may weaken regenerative ability. we test the hypothesis based on RNA-seq datasets for species with regenerative ability.

Expanded summary*: From hundreds years ago to now, scientists started to study regeneration, although we get some impressive progress, it is a complicated biological process and still remain a major challenge in biology. As we all know, from microorganisms to mammals, all species have regenerative ability, which is a capacity for regrowing or restoring the lost body part. Some species such as flatworms, which can regenerate a new individual from a small body fragment, have strong regenerative ability. On the contrary, other ones appear to have been greatly restricted. It has been reported that only a few genes (CAP-43, homeobox *msx*, Wnt, etc.) are related to regenerative ability (Fu et al., 1997; Liu et al., 2013; Akimenko et al., 1995), although many of genes may be involved in developing and resorting the lost body part during the regenerative process. Akimenko et al (1995) indicated expression level of homeobox *msx* strongly induced when regenerated the caudal fins in zebrafish, the level increased and reached a maximum between third and fifth days as the blastema formed, after 12 days, the fin became normal sizes. Raya et al (2014) also verified heart regeneration involved up-regulation of *msx* family genes (*msxB* and *msxC* genes included in this study). Moreover, they indicated the Notch pathway played a role when regenerating the heart of zebrafish. Furthermore, the evolution of genes with regenerative ability among the animals is still unclear. We hypothesize that there are master genes related to regeneration in animals, and losses of them or nosynonymous substitutes few amino acids, may weaken regenerative ability. we test the hypothesis based on RNA-seq datasets for species with regenerative ability.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-232

Comparative Transcriptomics and RiboSeq: Looking at De Novo Gene Emergence in Saccharomycotina

William Blevins [1 2,*] on behalf of Evolutionary Genomics Group, Mar Albà [1 2] on behalf of Evolutionary Genomics Group, Lucas Carey [2] on behalf of Single Cell Behavior Group

[1]GRIB, Hospital del Mar Research Institute, [2]DCEX, Universitat Pompeu Fabra, Barcelona, Spain

Abstract: In *de novo* gene emergence, a segment of non-coding DNA undergoes a series of changes which enables transcription, potentially leading to a new protein that could eventually acquire a novel function. Due to their recent origins, young *de novo* genes have no homology with other genes. Furthermore, *de novo* genes may not initially be under the same selective constraints as other genes. Dozens of *de novo* genes have recently been identified in many diverse species; however, the mechanisms leading to their appearance are not yet well understood. To study this phenomenon, we have performed deep RNA sequencing (RNA-Seq) and ribosome profiling (Ribo-Seq) experiments on 11 species of yeast from the phylum of Ascomycota in both rich media and oxidative stress conditions. These data have been used to classify the conservation of genes at different depths in the yeast phylogeny. Hundreds of genes in each species were novel (unannotated), and many were identified as putative *de novo* genes; these candidates were then tested for signals of translation using our Ribo-Seq data. We show that putative *de novo* genes have different properties relative to phylogenetically conserved genes. This comparative phylotranscriptomic analysis advances our understanding of *de novo* gene origins.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-248

Comparative analysis of angiosperm genomes reveals lineage-specific patterns of gene turnover and gene family evolution in Brassicaceae

Xuan Lin [1,*], Claudio Casola [1]

[1]Department of Ecosystem Science and Management, Texas A&M University, College Station, United States

Abstract: Gene duplication and gene loss are major sources of genetic variation between species. The abundance of genomes currently available from a variety of taxa allows to systematically address patterns of gene duplication and loss along phylogenies. Here, we utilize 856,618 genes (11,179 gene families) from 44 flowering plants available in the Phytozome database to identify gene duplication and gene loss events with a particular emphasis on Brassicaceae and the *Arabidopsis thaliana* lineage. Maximum-likelihood inferences of gene family size for each node of this phylogeny were obtained using the CAFE software on a subset of 9,815 families present in at least 20 species. A total of 1,220 gene families exhibited accelerated gene turnover (Viterbi algorithm, *P-value* <0.05). A Gene Ontology term enrichment analysis indicated that genes involved in general biological processes, particularly transcription and translation, are common among 3,556 *A. thaliana* genes from these families. A subset of 221 gene families had gone through significant expansion (122 families) or contraction (99 families) in *A. thaliana*. The 512 *A. thaliana* genes in families with significant size increase showed enrichment in metabolic pathways, including terpenoid biosynthesis, and transport of molecules, such as sucrose transporters. Most gene families with significant decrease in size in *A. thaliana* appeared to have been entirely lost in the mustard weed. Therefore, we searched for functional enrichment in 160 genes from *Populus trichocarpa* occurring in these gene families, of which only a handful were annotated under the 'response to stress' GO category. Comparison across a larger cohort of gene families that occurred in at least 4 of the 44 analyzed species showed lineage-specific patterns of gene turnover. For instance, we identified 685 lineage-specific gene families shared by 10 Brassicaceae species. Interestingly, no functional enrichment was observed among 847 *A. thaliana* genes from these families except for cell-cell signaling process (7 genes) and endomembrane cellular localization (205 genes). On the contrary, gene families lost in Brassicaceae but present in *P. trichocarpa* showed functional enrichment in a variety of processes, particularly response to stress, protein biosynthesis and gene expression regulation. Our analysis highlights a dynamic landscape of gene turnover both at the level of family (Brassicaceae) and single species (*A. thaliana*) lineages.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-230

Expansion of gene families and signatures of selection in the Australian marsupials

Will Nash [1,*], Wilfried Haerty [1]

[1]The Earlham Institute, Norwich, United Kingdom

Abstract: The Marsupials are thought to have diverged from other mammalian taxa around 160 million years ago. Since this time, Australian marsupial lineage has undergone a unique radiation, dominated by a lack of large predators, and a need to adapt to extreme heat and dryness and specific diets. Gene family expansion and contraction as long been shown to be a unique process often associated with evolutionary innovations and ecological adaptation. The recent availability of a high-quality genome for the koala (*Phascorlarctos cinereus*), in addition to those of three other marsupial species provides the opportunity to assess gene family dynamics within this unique lineage. In the koala lineage, we recover signatures of expansion in over 1,000 gene families. Such signatures were recovered in a range of immunoglobulin variable region families, and gene families associated with reproductive traits such as spermatogenesis. The largest expansion was found within the CYP2C subfamily, representing two independent expansions in koala including a total of 33 novel Cyp450 gene duplications. This is of particular interest as such Cyp450 proteins play essential roles in the metabolism of xenobiotic compounds such as the toxic turpenes found in great abundance in the koala's eucalyptus diet. An analysis of the conserved synteny of these genes also allowed their chromosomal placement, and showed the expansion within the marsupials to be of independent origin to a similar independent expansion found in rodents. Additionally, we analysed genes that are 1:1 orthologs across the tree of 9 species used here, for signatures of positive selection. In the koala, these genes with evidence of positive selection were enriched for gene ontology terms associated with growth, muscular migration, sexual reproduction, and various responses to stress. Within Australidelphia we found a range of pathways to be enriched for genes with evidence of positive selection, and of particular interest the thyroid hormone synthesis pathway, an organ previously described for its unique morphology among mammals.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI12

Variation and novelty in evolution: de novo genes arise and enable protein structural innovation

Victor Luria [1,*], Amir Karger [2], Anne O'Donnell-Luria [3], Taran Gujral [4], Bradley Olson [5], Wes Cain [6], Marc Kirschner [1]

[1]Systems Biology, [2]IT - Research Computing, Harvard Medical School, Boston, [3]Broad Institute, Harvard University, Cambridge, [4]Human Biology, Fred Hutchinson Cancer Research Center, Seattle, [5]Biology, Kansas State University, Manhattan, [6]Mathematics, Harvard University, Cambridge, United States

Abstract: Protein domains are autonomously folding units that carry the function of protein-coding genes by enabling interaction with other proteins or performing biochemical functions. How new protein domains are invented and how new protein-coding genes appear are major unsolved questions in evolutionary biology. New protein-coding genes are typically built using old genes (genes or genome duplication, domain shuffling, retrotransposition). In contrast, the existence of new genes that arise *de novo*, directly from genomic DNA, has long been ruled out. Recently *de novo* protein-coding genes were found in many eukaryotic genomes. Not using parts of ancient genes, *de novo* genes are the only ones that can encode genuinely new protein structures. *De novo* genes function and are generally expressed in neurons and male germline. We found an ancient orphan gene in humans that may have arisen *de novo* and showed it has important functions in neurons and skin. To understand how often *de novo* genes appear, what proteins they make and what functions they have, we take a three-pronged approach – mathematical, computational and experimental. To assess how many *de novo* genes can arise, we built a mathematical birth-and-death model based on gene and genome parameters and on dynamic factors such as mutation, recombination and selection. We found every genome should have many new genes, of which few are kept. We then identified thousands of candidate *de novo* genes in 20 eukaryotic genomes using phylostratigraphy and proteomics data. Using structural bioinformatics, we evaluated their predicted biophysical properties. Compared to ancient proteins, *de novo* genes encode proteins that are shorter, vulnerable to proteases, disordered, likely to bind other proteins yet less prone to aggregation, which is toxic. We started biophysical experiments comparing human *de novo* proteins to human ancient proteins, proteins of intermediate age, possible proteins from human genomic ORFs and random proteins. We will follow with functional experiments in neurons. Here we aim to determine the general properties of *de novo* genes in eukaryotic genomes. We find that many *de novo* genes arise and encode short but non-toxic proteins, and that few survive evolutionarily. The few survivors may encode genuinely new protein structures, suggesting genomes continually generate variation that enables new structures, new functions and is selected upon.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-233

Is Maternal Effect Dominant Embryonic Arrest (MEDEA) explained by DUF1703 mediated target immunity?

Catherine Rogers [1,*], Jeffrey Demuth [2]

[1]Biology, University of TX arlington, Joshua, [2]University of TX arlington, Arlington, United States

Abstract: Maternal Effect Dominant Embryonic Arrest (MEDEA) is a selfish genetic factor that is found in the flour beetle, *Tribolium castaneum*. MEDEA promotes its own frequency in populations by causing non-MEDEA offspring of heterozygous MEDEA mothers to be inviable. There are at least two MEDEA factors known to segregate among *T. castaneum* populations (M1 and M4). M1 is the best characterized, being composed of a composite Tc1-like transposon that carries a large (~21kb) insertion. Within the insertion there is only one intact ORF, called DUF1703. Recent computational work by Zhang et al (2016) revealed homology between parts of the M1 insertion and Crinkler (Crn) effector proteins, a phylogenetically widespread family of selfish elements commonly recruited to roles in inter- and intra-genomic conflict. Inside the Crn effector paralog found in *T. castaneum* is a protein named DUF1703. They propose that DUF1703 is the protein responsible for silencing the catalytic activity through "target immunity". Target immunity is a phenomenon found in some DNA transposons where they encode a protein to discourage re-insertion back into locations that already contain the transposon. Target immunity occurs when a protein is produced that binds to the target DNA and thereby displaces normal nucleolytic activity of the transposase. If target immmunity is the mechanism behind MEDEA offspring survival, we expect DUF1703 to be the protein in MEDEA positive progeny responsible for binding target DNA, thus protecting it from cleavage by the transposase. To test the target immunity hypothesis we first test whether DUF1703 is expressed in the necessary time and tissue to affect the MEDEA phenotype. Then we attempt to functionally validate target immunity by using RNAi to knock down DUF1703 expression.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-228

Functional analyses of transposable element-derived genes in Drosophila melanogaster

Diwash Jangam [1,*], Cedric Feschotte [2], Esther Betrán [1]

[1]Biology, University of Texas at Arlington, Arlington, [2]Department of Human Genetics, University of Utah, Salt Lake City, United States

Abstract: Several domesticated transposases from *PIF/Harbinger* DNA transposable elements (TEs) have been described in *Drosophila* by Betrán and Feschotte labs. Here, we present work aimed to elucidate the function of two *PIF/Harbinger* derived transposases named *Drosophila PIF like gene 1* (*DPLG1*) and *DPLG4*. GO enrichment analysis of *DPLG1* and *DPLG4* co-expression cluster, as defined by a recent high-resolution analysis of the developmental transcriptome of *D. melanogaster* by Graveley *et al.* 2011, shows these genes co-express with transcription factors and the study by Casola *et al.* shows that these genes have retained their HTH DNA binding domain but have lost their DDE catalytic triad. Thus we postulate that these proteins are domesticated for their ability to recognize and bind specific DNA sequences and are now transcription factors. Additionally, since TE proteins are able to recognize TE sequences, we further hypothesize that these genes may have been potentially domesticated to defend against transposition events.

Taking advantage of the UAS/GAL4 system, *DPLG1* was knocked down in all tissues using *Actin5c-GAL4* driver and a UASt-RNAi-*DPLG1* line, and we performed RNA-Seq analysis of the *DPLG1* knockdown (KD) testis. The testis of these KD males show increased activity of LINE TEs including all three telomeric elements. QRT-PCR analyses of these samples showed significant difference only in the most differentially expressed TEs. We have recently generated null mutants of *DPLG1,* in collaboration with Buszczak lab, and we will be testing if the effect in the *DPLG1* Knockout (KO) flies would be amplified from what we observe in the KD testis. *DPLG4* might also be involved in TE control. A study by Handler *et al.* showed that *DPLG4* scored positive in a large RNAi screen for new ovarian piRNA factors. In a screen of over 7000 genes, *DPLG4* was one of the 368 genes that scored positive for upregulation of gypsy-lacZ reporter in the ovarian somatic cells. We have evidence which shows that DPLG4 localizes in the nucleus of the female germline and thus, could potentially be interacting with DNA to regulate TE activities. We have generated a null mutant of *DPLG4,* also in collaboration with Buszczak lab, which we will be using to further investigate the role of *DPLG4* in TE regulation. Additionally, our data from the KO and KD experiments shows that *DPLG4* is important for viability, and possibly fertility in *D. melanogaster*.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-229

Successive gene duplications drive enrichment of dominance in ortholog models over evolutionary time

Daniel Jordan [1,*], Daniel Balick [2], Shamil Sunyaev [2 3], Ron Do [1]

[1]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, [2]Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, [3]Department of Biomedical Informatics, Harvard Medical School, Boston, United States

Abstract: The use of model organisms in the study of human biology relies on the assumption that the relevant biology is sufficiently conserved, and that the intervening evolutionary time has not produced enough innovation or functional divergence to disrupt the usefulness of the model. This consideration is particularly important for model organisms that are separated from human by very large spans of evolutionary time, such as the fruit fly *D. melanogaster*, the nematode *C. elegans*, and the yeast *S. cerevisiae*, all of which are commonly used as model systems for human biology. We investigated whether the relative prevalence of dominant and recessive alleles is conserved between humans and these model organisms, using our own novel method for predicting dominance and selection in human genes. This method is based on comparing the site frequency spectrum (SFS) of simulated genes of known dominance and selection with that of observed variation in approximately 35,000 Europeans in the Exome Aggregation Consortium (ExAC) dataset. Our predictions of dominance revealed a significant enrichment in dominant human genes among genes with known orthologs in fly, nematode, and yeast, with the magnitude of the enrichment increasing with increasing phylogenetic distance. We found the same signal in curated lists of dominant and recessive disease genes collected from previously published literature. We also found a reciprocal enrichment of fly genes annotated as dominant in FlyBase among genes with known orthologs in human, nematode, and yeast. Possibly explaining this phenomenon, we found that the number of duplicated genes as a fraction of orthologs increases with increasing phylogenetic distance from human. It has been previously observed that genes that have undergone duplication are more likely to be dominant, since genes that are sensitive to gene dosage are likely to be maintained as duplicates. Hence, dominant genes become enriched over evolutionary time due to having systematically larger numbers of orthologs in more distant organisms.
To demonstrate the relevance of this phenomenon to model organism research, we collected literature-derived lists of human disease genes studied in yeast and fly. We found statistically significant enrichments of predicted dominant genes in these lists. The literature may therefore have an overall bias towards investigating traits that are under dominant selection, while those under recessive selection may be systematically understudied.

Expanded summary*: While many methods exist to predict the action of selection on protein-coding variants or genes, most of these are implicitly focused on detecting selection on heterozygous alleles. This includes both methods that rely on conservation like SIFT, PolyPhen-2, GERP++, and phyloP, and methods that rely on population variation like RVIS and pLI. This situation persists even though it is well known that some alleles experience different selection in heterozygous and homozygous; in fact, this is one of the oldest and best-established results in genetics. This is because recessive selection is often very difficult to detect, mostly due to the fact that most statistics that respond to recessive variation are heavily confounded by demography.

We have developed a novel method to predict the dominance and selection of human protein-coding genes. We address the problem of confounding by demography by using the Exome Aggregation Consortium (ExAC)'s dataset of approximately 35,000 European exomes, which have a known and homogeneous demography. We simulate genes of fixed selection coefficient *s* and dominance coefficient *h* through this demography and match these simulated genes to the observed genes from ExAC to produce an joint estimate of dominance and selection for each gene. Unlike previous methods, this method is capable of distinguishing recessive selection from both dominant selection and neutrality.

As an application of our method to biological discovery, we set out to test the conservation of dominance coefficients over evolutionary time. We discovered that vertebrate orthologs of human genes have largely the same makeup of dominance as the human genome overall. However, this result does not hold for non-vertebrate model organisms including *D. melanogaster, C. elegans,* and *S. cerevisiae,* all of which are significantly enriched for orthologs of genes predicted to be under dominant selection in humans. We

discovered that this signal coincides with and appears to be driven by an increased number of gene duplications between these species and human. In fact, the fraction of these organisms' genomes that have one-to-many or many-to-many orthology relationships with human genes directly predicts the level of enrichment of dominant genes. This seems to be explained by a previously observed result that duplicated genes are more likely to be dominant, since duplications confer a selective advantage where genes are sensitive to dosage. Thus, orthologs of dominant genes proliferate over evolutionary time due to gene duplication events, while orthologs of recessive genes do not. This has implications for the use of model organisms and comparative genomics to study human disease, as researchers are more likely to find orthologs for dominant human disease than recessive. This suggests that the bias against recessive selection may not be limited to algorithms, but may also extend to model organism research. We demonstrated this bias by collecting lists of genes implicated in human disease that are studied in various model organisms, including fly, yeast, and zebrafish. As expected from the trend we observed genome-wide, we find a statistically significant enrichment of genes under dominant selection in the yeast and fly lists, and a sub-significant enrichment in the same direction in the zebrafish list. An even larger portion of the biological literature than expected may therefore show this bias against genes and phenotypes under recessive selection. This further highlights the necessity of unbiased ways to detect recessive selection, such as our method.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-250

Comprehensive reconstruction of gene evolutionary histories reveals spatio-temporal biases in the dynamics of horizontal gene transfer

Alexander Esin [1 2,*], Tobias Warnecke [1 2]

[1]Imperial College London, [2]MRC London Institute Of Medical Sciences (LMS), London, United Kingdom

Abstract: Horizontal gene transfer (HGT) is a major driver of microbial evolutionary change and has been implicated in the origin of distinct prokaryotic lineages. Despite its abundance, it is clear that the process is not random – HGT is known to be biased by the function of the gene product and phylogenetic distance between donor and recipient. Less clear is the presence of biases over evolutionary time and genome space. In part this is because of the difficulty of accurately and comprehensively determining HGT events over a broad evolutionary timescale. Here, we investigate the temporal (continuous vs. punctual gain) and spatial (genomic location) dynamics of HGT, as well as the interplay between the two. We apply a rigorous phylogenetic approach and model HGTs in the context of alternative evolutionary scenarios to identify transfer events from across the tree of life into a single bacterial genus, *Geobacillus*. We find that the number of transfers along a particular branch strongly correlates with the length of that branch, supporting a model of continuous HGT acquisition in *Geobacillus*. We also identify a non-random spatial distribution of horizontally acquired genes characterized by: a) enrichment in HGTs near the terminus, b) substantial clusters of HGT-derived genes near the origin of replication, and c) an unexpected depletion of HGTs midway along the replichores. Curiously, this pattern is symmetric along the origin-terminus axis and spatial biases remain remarkably consistent between species separated by over 100 million years.

Expanded summary*:

Horizontal gene transfer (HGT), largely in the form of transformation, has been harnessed as a laboratory tool for genetic modification for over half a century. The large role of HGT in the evolution of prokaryotes, on the other hand, has only recently become apparent along with some global "rules" or recurring patterns that characterize successful transfers. For example, genes that can work alone, such as enzymes, are more frequently transferred than members of large interdependent protein complexes. Similarly, there has been a lot of progress in elucidating donor-recipient relationships, highlighting the importance of gene sharing highways. In comparison to the "what" and "where from", we know much less about the "when" (i.e. timing of HGT acquisition during lineage evolution) and the "where" (i.e what governs the location of gene integration into the recipient genome?).

Regarding the "when", Nelson-Sathi et al. (2015) recently proposed that births of major archaeal clades were marked by punctuated bursts of HGT at their roots, defining their colonization of novel niches. Groussin et al. (2015) contested that the data was equally consistent with transfers being accrued throughout lineage evolution – a continuous model of HGT acquisition. Regarding the "where", genomic islands (GIs) have been implicated as hotspots of HGT ingression, but whether HGTs acquisitions are restricted to GIs, how acquisition events are constrained by genome architecture, and how transfers can escape into the rest of the genome largely remains to be elucidated.

In my project, I have been studying the "where" and "when" (as well as the "what") of HGT into a single defined clade, *Geobacillus*. In part, we have focused on a single clade because a major challenge in HGT inference is ruling out alternative evolutionary scenarios: duplication and loss, which – if done comprehensively – is computationally intensive. To study the timing and the topological biases of HGT we employ a rigorous phylogenetic analysis to detect HGTs into the thermophilic *Geobacillus* genus, from over 4000 species spanning the tree of life. Our analysis is based on a reconciliation method that explicitly considers transfers, duplications, and losses in deducing the most parsimonious resolution to phylogenetic incongruence between gene and species trees.

Our conservative set of HGTs into *Geobacillus*, accounting for ~20% of the pangenome, is functionally enriched in metabolic genes – in line with previous works. Regarding the "when", we find that the number of transfers correlates very strongly ($r = 0.79$) with branch length, a proxy for evolutionary time. This correlation is strongly suggestive of a continuous model of HGT acquisition in *Geobacillus*, though we cannot rule out the contribution of small HGT bursts over the course of lineage evolution. Regarding the

"where", we observe enrichment in HGTs around the terminus and, surprisingly also near the origin of replication, while we observe an unexpected and strong depletion midway along the replichores. In all species, the observed enrichment/depletion pattern is strikingly symmetric along the origin-terminus axis, further highlighting the apparent non-randomness of HGT distribution in *Geobacillus* genomes and suggesting interplay between topological bias of HGT events and adaptive genome architecture. Combining "when" and "where", we find that the genomic positions of these transfer events are remarkably consistent along different branches.

Our characterization of horizontal transfer into *Geobacillus* contributes to our growing understanding the evolutionary history of gene flow and its impact on adaptation. It also provides important pointers for rule-based genome engineering, especially regarding the recurrent topological biases we unveil, which we suggest are likely to be a more general feature of HGT dynamics across prokaryotes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI7

Stop codon readthrough errors purge deleterious cryptic sequences, facilitating the later co-option of non-coding sequences into coding

Luke Kosinski [1,*], Joanna Masel [2]

[1]Molecular and Cellular Biology, [2]Ecology and Evolutionary Biology, University of Arizona, Tucson, United States

Abstract: *De novo* genes originate from ancestrally non-coding DNA and are a source of genetic and phenotypic novelty, but how they emerge is unclear. The "pre-adapting selection" hypothesis states that this emergence is facilitated by the fact that non-coding sequences may be translated in error, creating polypeptides that may either interfere with existing cellular machinery, making them deleterious, or may have a minimal impact, making them benign. Selection favoring more benign sequences is expected to be strongest when expression is higher.

To test for pre-adapting selection acting on non-coding sequences lying beyond the stop codons (i.e. readthrough products) of *Saccharomyces cerevisiae* genes, we looked for an association between the amount of readthrough and whether the readthrough product was benign. Increased expression of readthrough products was assumed to correlate with protein abundance (from PaxDB) and with the existence of ribosome profiling hits in a gene's 3′ untranslated region (UTR). Benign status was assumed to correlate with high intrinsic disorder (ISD, measured using IUPred), low aggregation propensity (measured using Waltz), and short length. Across this variety of metrics, results supported the prediction of pre-adapting selection. Moreover, the readthrough product itself, rather than the neighboring C-terminal amino acids, is an important driver of selection. These findings indicate that selection acts on 3′ UTRs in *S. cerevisiae* to purge potentially deleterious variants, acting more strongly in genes which experience more errors.

Expanded summary*: How do new genes originate? The classic answer is that existing genes beget new genes through the process of gene duplication. But this answer is imperfect: it implies that all genes have homologs, but genes with no identifiable homologs have been found. It also fails to explain how the first genes arose. Recent evidence has shown that genes may emerge *de novo* from non-coding sequences. This finding shocked researchers because genes are thought to develop over long periods of time through continual tweaking and editing by natural selection. The idea that a polypeptide from a non-coding sequence, here called a "potential polypeptide," without a history of exposure to evolution, could suddenly become a gene was considered absurd. However, if potential polypeptides are exposed to selection, such as through molecular errors, deleterious variants could be weeded out and benign variants would remain, providing a history of exposure to evolution. This idea is encapsulated in a theory called "pre-adapting selection," with the specific reasoning that the more errors occur that translate a potential polypeptide, the more it gets exposed to selection, and the more likely it is to be benign. The goal of the present project is to find evidence of pre-adapting selection.

The obvious way to show a link between pre-adapting selection and *de novo* genes would be to examine intergenic open reading frames (ORFs) and the correlation between their expression through error and whether or not they are benign. However, it is difficult to quantify translation for sequences that may be translated only rarely or not at all. Existing ribosome profiling datasets are currently up to this task only in *Saccharomyces cerevisiae*, which has short intergenic regions and thus few intergenic ORFs.

Pending new and improved datasets in a broader range of species, "proof of concept" for the pre-adapting selection hypothesis can be established using not complete intergenic ORFs, but instead potential polypeptides that get appended to an existing protein due to readthrough errors. During a readthrough error, a stop codon is bypassed by the ribosome and the 3′ untranslated region (UTR) sequence is translated through to the next in-frame stop codon. According to the pre-adapting selection hypothesis, the more frequently readthrough errors occur, the more likely the readthrough product is to be benign.

To test for pre-adapting selection acting on readthrough products of *S. cerevisiae* genes, we looked for an association between the amount of readthrough and whether the readthrough product was benign. Increased expression of readthrough products was assumed to correlate with protein abundance (from PaxDB) and with the existence of ribosome profiling hits in a gene's 3′ untranslated region (UTR). Benign status was assumed to correlate with high intrinsic disorder (ISD, measured using IUPred), low aggregation propensity (measured using Waltz), and short length. Across this variety of metrics, results supported the prediction of pre-adapting selection. Moreover, the readthrough product itself, rather than the neighboring C-terminal amino acids, is an important driver of selection.

These findings indicate that selection acts on 3′ UTRs in S. cerevisiae to purge potentially deleterious variants, acting more strongly in genes which experience more errors.
These results show that pre-adapting selection occurs in yeast in response to readthrough errors. This means that a mechanism to make potential polypeptides more benign, and thus more amenable to becoming a *de novo* gene, is not only plausible but has some evidence of its operation in biological systems. The next step is to examine intergenic ORFs; this is quickly becoming more feasible as new datasets emerge, and has the potential to strengthen the link between the pre-adapting selection hypothesis and *de novo* genes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-251

Evolution reorganized an innate antiviral immune signaling network in early animals

Charles Pugh*, Oralia Kolaczkowski [1], Bryan Kolaczkowski [1]

[1]University of Florida, Gainesville, United States

Abstract:

The ability to sense and respond to viral infection is paramount to the survival of organisms of all kinds. In contrast to the considerable effort spent on understanding the function and development of vertebrate immune systems, relatively little is known about antiviral immune function in very early animals. Understanding how innate immune signaling works in early animals can help elucidate how long-term host-pathogen coevolution may drive molecular-functional innovations. Here we characterize the evolutionary histories of the RIG-like receptors (RLRs)—a family of innate immune proteins that recognize viral RNA—and their immediate downstream signaling partner, IPS1. We find that at least two RIG-like receptors were present in the model Cnidarian, Nematostella vectensis, although the IPS1 signaling adaptor did not evolve until much later in deuterostomes. This finding poses the conundrum, how did early eumetazoan RLRs signal before their adapter evolved? We identify a putative signaling adapter from N. vectensis transcriptomic data, clone this adapter and the two N. vectensis RLRs from cDNA, and demonstrate the potential for direct protein-protein interactions between N. vectensis RLRs and this putative signaling partner, which we call "NIPS1." We use ancestral sequence reconstruction, structural modeling and molecular kinetics experiments to demonstrate how historical substitutions in RIG-like receptor signaling domains were responsible for an evolutionary shift in the RLR signaling network, from early eumetazoan interactions with NIPS1 to a preference for IPS1 in deuterostomes. This work demonstrates the mechanisms by which pathogen-driven sequence evolution can rewire signaling networks critical for organism survival.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI13

Problems and solutions in homology search-based de novo gene identification: toward a computational control for undetected homologs

Caroline Weisman [1,*], Andrew Murray [2], Sean Eddy [2]

[1]Program in Biophysics, [2]Molecular and Cellular Biology, Harvard University, Cambridge, United States

Abstract: Genes that have emerged from previously non-genic sequence ("de novo genes") represent a recently recognized form of evolutionary innovation. Compared to more canonical processes of gene addition, such as horizontal transfer or duplication, de novo emergence is poorly understood. The mechanisms underlying de novo gene emergence, the rate at which they emerge, and the biological roles of these de novo genes are among the questions that we would like to answer in order to more fully understand their evolutionary importance.

A first step in answering these questions is to identify a set of genes that are good candidates for having originated de novo. One common method for such identification is to use results of homology search algorithms such as BLAST: if a gene lacks an identifiable homolog outside of a narrow clade, it is inferred to have originated de novo along the founding lineage. However, recent work has questioned the susceptibility of this method to errors wherein a homolog is actually present outside the clade but is not detected, resulting in the gene being falsely called de novo.

Here, we present a simulation-based tool to determine how susceptible a gene is to this form of erroneous de novo classification: the probability that, even if its homolog were present outside the clade, it would not be detected. Previous work has used simulation-based approaches to address the issue of such erroneous de novo calls, but has also demonstrated the sensitivity of results to the parameters of the simulation, which have been the subject of debate. We thus also present the first method to quantitatively assess the accuracy of the simulations used in such a tool. Taken together, our work decreases the risk of falsely calling genes with present but undetected homologs as de novo, and thus constitutes a method for more robust homology search-based de novo gene identification.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI8

Foldability of a de novo evolved protein

Matthew Cordes [1,*], Dixie Bungard [1], Jacob Copple [1], Jimmy Chhun, Vlad Kumirov [1], Jing Yan [2], Scott Foy [1], Joanna Masel [1], Vicki Wysocki [2]

[1]University of Arizona, Tucson, [2]Ohio State University, Columbus, United States

Abstract: The de novo evolution of protein-coding genes from noncoding DNA is emerging as a source of molecular innovation. It is unclear whether early-stage de novo proteins tend to be intrinsically disordered proteins or are able to fold into compact, specific structures like those that underlie the function of most highly evolved globular proteins. Here we show that Bsc4, a natural de novo protein encoded by a whole gene that evolved very recently from noncoding DNA in the yeast *S. cerevisiae*, folds spontaneously into a stable, but not entirely specific, three-dimensional structure. Recombinant Bsc4 expressed in *E. coli* folds into a soluble, compact structure with predominantly beta-sheet secondary structure content as judged by far ultraviolet circular dichroism. Tryptophan fluorescence suggests the presence of tertiary interactions, but a weak near ultraviolet circular dichroism spectrum suggests that it lacks a completely specific tertiary fold. Nonetheless, folded Bsc4 undergoes cooperative, reversible thermal denaturation and cooperative chemical denaturation by guanidine. Bsc4 lacks a specific quaternary structure, but exists as a fairly narrow distribution of small oligomeric states as judged by size exclusion chromatography and mass spectrometry. Bsc4 oligomers also bind thioflavin T and Congo Red, suggestive of similarity to pre-amyloid oligomers. Overall, the combination of "native-like" and "molten" or pre-amyloid properties suggest that Bsc4 has a rough, rudimentary folding pattern; such "proto-folds" could act as functional scaffolds and may provide a way for new folded, globular proteins to emerge de novo by gradual mechanisms.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI3

Molecular innovation in ciliates with complex genome rearrangments

Rafik Neme [1,*], Leslie Beh [1 2], Robert Bakaric [3], Laura Landweber [1 4]

[1]Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY, [2]Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, United States, [3]Exaltum Ltd, Zagreb, Croatia, [4]Biological Sciences, Columbia University, New York, NY, United States

Abstract: Ciliates, unicellular eukaryotes with two types of nuclei, possess the remarkable ability to rearrange their genomes in processes that involve the selective deletion, ligation and reorganization of genetic information from a sexual genome into a somatic genome. These arrangements can be simple or complex, depending on whether fragments of DNA are joined in the same order in the final product, or whether the rearrangements require translocation or inversion. This type of genomic architecture provides abundant plasticity and increased potential for innovation relative to other eukaryotic lineages.
We are interested in the evolutionary steps leading to this type of genomic architecture, and the influence it has had on evolutionary innovation. For example, we have previously reported that alternative DNA processing provides yet another mechanism for the emergence of new genes in the *Oxytricha* lineage (GBE 7: 2859-2870).
Now we have completely sequenced and assembled the germline genome for *Tetmemena pustulata*, a ciliate distantly related to yet in the same lineage as *Oxytricha trifallax.* Both genomes undergo extensive descrambling in the differentiation of a zygotic to somatic genome. Comparative genomic analysis of *Oxytricha trifallax* and *Tetmemena pustulata* is yielding insights into the evolutionary origin of lineage-specific genes, and how innovation in these ciliates both relates to and differs from other eukaryotes.

Expanded summary*: Molecular innovation profits from the organization of biological information. Most organisms obtain novelty from mechanisms such gene duplication, gene fusions, alternative splicing, or *de novo*.
In ciliates, it can be assumed that all such mechanisms are generally active. Furthermore, it has been shown that alternative DNA processing is able to produce new genes (GBE 7: 2859-2870), and as such constitutes a ciliate-specific innovation that produces further innovation. This is likely to have contributed greatly to the molecular and functional diversification on stichotrich lineages.
We compare the germline genomes of *Tetmemena pustulata* and *Oxitricha trifallax* (fully sequenced and assembled in our group), and inquire about the patterns of rearrangements and functional diversification of genes which have been acquired by the stichotrich lineage after they developed their massive genome rearrangement programs. We also explore transposon-derived elements, rates of rearrangement errors during development, and expression patterns to produce a model of how innovation takes place in a genome able to withstand the fragmentation and scrambling of its genome.
In addition to this, we undertake phylostratigraphic analyses of various available ciliate species, to highlight how innovation could have been accelerated in the lineages with complex genome rearrangements. This is a unique system to explore how a genomic-level innovation affects the general course of the genome evolution and innovation in a lineage.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-237

Characterization of Orphan Gene CRE23514 in Caenorhabditis remanei

Aubrey Kent [1,*], Rose Reynolds [1]

[1]Biology, William Jewell College, Liberty, United States

Abstract: As a special case of the evolution of new genes, so-called "orphan genes," for which there are no known homologs in other species, may present a unique opportunity to characterize the general functions and evolutionary relevance of de novo genes arising from noncoding genomic sequence. Here I propose to identify the function(s) of CRE23514, an orphan gene unique to the non-parasitic soil roundworm species *Caenorhabditis remanei*. This gene produces mature mRNA, but has no previously identified function or homologs. CRE23514 is upregulated over 8-fold in a population of *C. remanei* artificially selected to withstand extreme levels of acute heat compared to the ancestral *C. remanei* population from which it was evolved. I propose to characterize the function(s) of CRE23514 by examining heat shock resistance ability and other organismal phenotypes in *C. remanei* worms with no expression, and worms with high expression, of CRE23514. Toward this end, I have designed a CRISPR/Cas-9 construct to remove the wild type CRE23514 coding region. My future studies will involve phenotyping the genome-edited CRE23514-null mutants, the design and creation of a multiple-copy CRE23514 expression vector, and the phenotyping of transgenic *C. remanei* worms overexpressing CRE23514.

Statement: I am currently a Junior at William Jewell College in Liberty, Missouri, and am majoring in biology. William Jewell is a small liberal arts college with a strong biology curriculum that emphasizes research. Each biology major is required to complete four semesters of individual research. Research requirements include writing a research proposal, present the proposal in poster format, several oral presentations in front of the biology department, and a written thesis. Although William Jewell does have a strong biology program, it is also limited on outside opportunities due to funding. Such financial restrictions reduce the possibility for many students to attend conferences such as SMBE's. I am also a non-traditional student and mother. My partner and I have four children, which presents responsibilities many of my educational peers do not have. These responsibilities make attending such conferences even more difficult for me to afford. The opportunity to attend the SMBE conference will be vital in aiding me to develop a direction for my educational and career path. I am very interested in the genetic basis of evolution, but have yet to find the right direction for me within the field. Having the opportunity to explore various topics at the SMBE conference will help me find a focus for my future as well provide an opportunity to make connections in the field. Networking will allow me to explore various topics and areas of the world that I may otherwise not be exposed to.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POB-415

GENETIC SCREENING OF THE COAGULATION PATHWAY GENES USING THE THROMBOSCAN TARGETED SEQUENCING PANEL

Hani ALHADRAMI [1 2,*], Ashraf Dalol [3]

[1]Center of Innovation in Personalized Medicine, King Fahd Medical Research Center, King Abdulaziz University, [2]Medical Laboratory Technology, King Abdulaziz University, Jeddah, [3]Center of Innovation in Personalized Medicine, King Fahad Medical Research Center , Saudi Arabia

Poster: Background

Thrombophilia is a condition where the blood has an increased tendency to clot. Blood clots can generate diseases such as deep vein thrombosis and pulmonary embolism. Thrombophilia can be acquired (due to autoimmune diseases, pregnancy, hormone therapy, malignancy, myeloproliferative disorders, postsurgical state and nephrotic syndrome) or inherited (due to genetics predisposition). Inherited thrombophilia is a result of DNA mutation in genes responsible for the production of blood clotting proteins (coagulation system). Antithrombin, protein S and protein C are the most important inhibitors of the blood coagulation system, as most of thrombotic patients have inherited deficiency of one of these proteins. While inherited thrombophilia can be caused by a number of mutations, the most common ones are factor V Leiden (FVL) and prothrombin (factor II). Factor V Leiden mutation is a single nucleotide point mutation (SNP) located at position number 506 and alters amino acid arginine to glutamine in FV gene. Prothrombin (factor II) is the precursor to thrombin, which is essential in the coagulation cascade and located on chromosome 11p11-q12. Prothrombin G20210A mutation (factor II mutation) is a SNP located at position 20210 and changes amino acid guanine to adenine in the prothrombin gene. This mutation is associated with high levels of prothrombin and was reported to increase the risk of thrombosis almost three fold. Patients with high levels of other procoagulants such as factors VIII, IX, XI, VII, fibrinogen, and Von Willebrand factor (VWF) are also at high risk of thrombosis.

Materials and methods

Next Generation Sequencing allows high throughput DNA sequencing and mutation detection at a low cost and high turnover. This in turns has a major influence in both clinical care and understanding susceptibility to thrombophilia. Therefore, we have designed the Thromboscan panel which will allow the simultaneous screening of 23 coagulation genes using the Ampliseq™ technology.

Results

In this study, a screening panel of 23 coagulation genes has been developed, optimized and tested for the early diagnosis of thrombophilia using the cutting edge technology of next generation sequencing. The results confirmed 99.26% coverage of the targeted genes with 430 amplicons with sizes ranges between 125-275 bp generating 81.56 kb of DNA sequence. We have demonstrated that this panel can be used on DNA extracted from peripheral blood or saliva.

Conclusion

The availability of this panel will help increase our understanding of genetic susceptibility to thrombophilia and other aberrant thrombotic events.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-M12

Cell Biology of Microproteins Encoded by smORFs

Alan Saghatelian*

Abstract: Recent advances in genomics and proteomics have identified hundreds to thousands of non-annotated protein-coding genes of less than 100 codons (small ORFs or smORFs), simultaneously revealing a blind spot in traditional gene-finding algorithms. smORFs encode miniproteins that are detectable in cells and tissues. Several of these smORFs/miniproteins are evolutionarily conserved, which indicated biological functions. Functional proteomics revealed that miniproteins partake in a range of cell biology indicating the value of mining the smORFeome for additional functional genes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-234

Sexual Dimorphism and Retinal Mosaic Diversification Following Gene Duplication in Butterflies

Adriana Briscoe*, Kyle McCulloch [1]

[1]Ecology and Evolutionary Biology, UC Irvine, Irvine, United States

Abstract: Animal lineages have repeatedly expanded and diversified the opsin-based photoreceptors in their eyes which underlie color vision. Despite an abundance of photoreceptor diversity in some animals, the selective pressures giving rise to new photoreceptors and their spectral tuning remain mostly obscure. In the butterfly *Heliconius erato*, both sexes express an adaptively evolving violet-sensitive receptor in their eyes together with a UV-yellow pigment on their wings; traits which evolved at the base of the genus *Heliconius*. Female *H. erato* eyes also express an ancestral ultraviolet-sensitive (UV) receptor. To understand the selective pressures giving rise to the *H. erato* eye we asked: 1) How did this sexually dimorphic eye evolve, and 2) What potential benefit is conferred by having a violet receptor? We first compared short-wavelength opsin expression patterns in both sexes in 23 species, using immunostaining and RNA-Seq. We identified six unique retinal mosaics and three distinct forms of sexual dimorphism based on ommatidial types within the genus *Heliconius*. Phylogenetic analysis revealed independent losses of opsin expression, pseudogenization events, and parallel evolution. Opsin expression patterns are hyperdiverse within *Heliconius*. Our observations give insights into the selective pressures underlying the origins of new visual receptors.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-235

Carbapenemase Evolution

Sona Garsevanyan [1,*], Miriam Barlow [1]

[1]School of Natural Sciences, University of California, Merced, Merced, United States

Abstract: Carbapenem resistant Enterobacteriaceae frequently cause urinary tract infections in elderly individuals. Carbapememases are responsible for the phenotype. To further our understanding of how the genes encoding carbapenemases have evolved in recent years and what evolutionary events have enabled them to become a formidable health threat we reconstructed phylogenies of NDM, VIM, KPC and Bla-B beta-lactamases. We identified numerous amino acids substitutions that appear to be functionally important. We also found evidence for homologous recombination contributing to their evolution. Carbapenemases appear to have evolved recently, likely in response to the use of antibiotics.

Expanded summary*: Carbapenemases pose one of the biggest threats as they confer resistance. Carbapenemases are continuously identified worldwide and numerous genes have been reported. Carbepenemases have been described as chromosomal and species specific, nonetheless mobile genetic elements have led to interspecies spread which has caused a global issue

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-236

Retroposition as a source of antisense long non-coding RNAs with possible regulatory functions

Oleksii Bryzghalov [1,*], Michal Szczesniak [1], Izabela Makalowska [1]

[1]Department of Integrative Genomics, Institute of Antropology, Adam Mickiewicz University in Poznan, Poznan, Poland

Abstract: Long noncoding RNAs represent a large and diverse class of transcribed RNA molecules with a length of more than 200 nucleotides that do not encode proteins. Based on recent studies they appear to regulate gene expression through a diverse group of mechanisms. Some of lncRNAs originate from retrocopies, intronless copies of the so-called parental genes originated in the process of retroposition. Part of them are transcribed from the antisense strand and therefore, are expected to have functions other than the corresponding retrocopies or their parental genes. We found 58 lncRNAs that were transcribed in antisense to 35 human retroposition derived copies of protein-coding genes. These lncRNAs share sequence similarity with the corresponding parental genes but in the antisense orientation, meaning they have the potential to interact with each other and to form RNA:RNA duplexes. The subsequent analysis of the RNA:RNA duplexes revealed 10 lncRNAs with potential regulatory roles exerted on their parental genes, which include stability control, pre-mRNA and mRNA processing. Our findings suggest that retroposition-derived, antisense lncRNAs might affect the expression and processing of parental genes in a number of ways. This statement is supported by the in silico base-pairing of the RNA molecules, followed by computational function assignment, co-expression data and, occasionally, correlation of expression and evolutionary conservation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-231

Putative regulatory roles of retrocopies nested in other genes

Magdalena Kubiak [1,*], Elżbieta Wanowska [1], Wojciech Rosikiewicz [1], Izabela Makałowska [1]

[1]Department of Integrative Genomics, Institute of Anthropology, Faculty of Biology, Adam Mickiewicz University in Poznan, Poznan, Poland

Abstract:

Retrocopies of protein-coding genes are duplicates created by reverse transcription of mRNA and reintegration of cDNA into a new genomic localization. A great number of reports show that many of retrocopies are transcriptionally active and therefore they have potential to play various molecular roles. Retrocopies were found to encode proteins as well as non-coding RNAs involved in different expression regulatory pathways.

In our research, we analyzed retrocopies localized in exons and/or introns of other genes and investigated their possible functions. Basing on human retrocopy repertoire from RetrogeneDB2 and Ensembl genes annotations, we identify 2139 retrocopies overlapping 2071 protein coding and non-coding genes. Out of them, 43 are integrated, entirely or partly, into exons of protein coding transcripts, 203 into non-coding transcripts and 14 into both types. Among them 169 are incorporated into exons in antisense orientation in comparison to their parental genes. Further bioinformatic analyses allowed us to identify 19 candidates that could potentially act as trans-NATs regulating expression of their parental genes. We were also interested in expressed intronic retrocopies and possibility of transcriptional interference between them and host genes. For this purpose, we searched retrocopies localized in intron of other gene, downstream of at least one shorter gene isoform. We found 58 retrocopies meeting these requirements. To further investigate their putative regulatory role, we deleted the chosen retrocopy, retro_hsap_4044, from its host gene in CRISPR/Cas9 experiment. In a result, we observed that after retrocopy deletion one of the shorter transcripts disappeared, what suggests that retrocopy transcription may play a role in the regulation of host gene's isoforms expression.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI14

De novo gene evolution: How do we transition from non-coding to coding?

Jorge Ruiz-Orera, José Luis Villanueva-Cañas, William R. Blevins, Mar Albà*

Abstract: P { margin-bottom: 0.08in; }

***De novo* gene evolution: How do we transition from non-coding to coding?**

Jorge Ruiz-Orera[1], José Luis Villanueva-Cañas[2], William R. Blevins[1], M.Mar Albà[1,3]

[1]Evolutionary Genomics Group, Research Programme on Biomedical Informatics, Hospital del Mar Research Institute, Universitat Pompeu Fabra, Barcelona. [2]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona. [3]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona.

Recent years have witnessed the discovery of protein–coding genes which appear to have evolved *de novo* from previously non-coding sequences. This has changed the long-standing view that coding sequences can only evolve from other coding sequences. However, there are still many open questions regarding how new protein-coding sequences can arise from non-genic DNA.

Two prerequisites for the birth of a new functional protein-coding gene are that the corresponding DNA fragment is transcribed and that it is also translated. Transcription is known to be pervasive in the genome, producing a large number of transcripts that do not correspond to conserved protein-coding genes, and which are usually annotated as long non-coding RNAs (lncRNA). Recently, sequencing of ribosome protected fragments (Ribo-Seq) has provided evidence that many of these transcripts actually translate small proteins. We have used mouse non-synonymous and synonymous variation data to estimate the strength of purifying selection acting on the translated open reading frames (ORFs). Whereas a subset of the lncRNAs are likely to actually be true protein-coding genes (and thus previously misclassified), the bulk of lncRNAs code for proteins which show variation patterns consistent with neutral evolution. We also show that the ORFs that have a more favorable, coding-like, sequence composition are more likely to be translated than other ORFs in lncRNAs. This study provides strong evidence that there is a large and ever-changing reservoir of lowly abundant proteins; some of these peptides may become useful and act as seeds for *de novo* gene evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-241

Mechanism and impact of de novo gene emergence in yeasts

Nikolaos Vakirlis [1,*], Alex Hebert [2], Dana Opulente [3], Chris Hittinger [3], Gilles Fischer [4 5], Joshua Coon [2], Ingrid Lafontaine [6]

[1]Department of Genetics, Trinity College, the University of Dublin, Dublin, Ireland, [2]Department of Biomolecular Chemistry, University of Wisconsin - Madison, Madison, United States, [3]Laboratory of Genetics, University of Wisconsin - Madison, Madison, United States, [4]LCQB, IBPS, UPMC, [5]CNRS, [6]IBPC, UPMC, Paris, France

Abstract: How new genes and new protein functions arise is a fundamental question in evolution. The creation of novel functions using existing genes as raw material (gene duplication, gene fusion etc.) is well established. Nonetheless, the discovery of de novo gene emergence from previously non-coding sequences proved that novel genes can also evolve "from scratch". Still, the underlying molecular mechanisms as well as the overall impact of de novo emergence on genome evolution are not entirely understood.
We developed a comprehensive methodology to achieve reliable, genome-wide de novo gene identification and applied it to the yeast genera *Saccharomyces sensu stricto* and *Lachancea*. The rate of emergence is steady within each genus but varies 10-fold between the 2 genera. Our results suggest that de novo genes have originated from non-coding regions that are significantly more GC-rich compared to the genome average. De novo genes were predominantly found divergently oriented relative to their neighbors, suggesting that their emergence was likely driven by divergent transcription from bidirectional promoters. In *Saccharomyces* de novo genes were found associated to recombination hotspots. These findings lead us to propose that in yeasts, the combination of 1) high GC% sequences around bidirectional promoters, a result of GC-biased gene conversion following Double Strand Breaks, and 2) constant divergent transcription of these same regions, provide conditions that favor de novo gene emergence by generating transcripts with lower probability for AT-rich stop codons. High GC% in these transcripts could also lead to (or simply indicate) higher transcriptional or translational efficiency.

Expanded summary*: How new genes and new protein functions arise is a fundamental question in evolution. New genes are essential to the evolution of novel phenotypes, to adaptation to new environments and to the process of speciation. Once considered so improbable as to be impossible, the origin of new genes de novo from previously non-coding sequences has now been demonstrated in every eukaryotic lineage studied, and these new genes have become integrated into central cellular functions [1–13]. New genes in general are linked to innovation [12-14] and de novo genes, in particular, may allow for innovation in a short evolutionary time-frame [13]. A good example of this is the yeast-specific de novo gene *MDF1* which, following its origination, rapidly acquired a double functionality, immediately advantageous to the cell [9,15]. In Drosophila, de novo genes have been consistently found more expressed in testis and have been thought to play a role in male reproductive processes [16–19]. In human, de novo genes have been found associated to cancer and Alzheimer's disease [1,2,5,20–22]. Several important questions in the study of de novo genes are still wide open [23]. The rate and the overall impact of de novo emergence remain largely unknown. Similarly, the genomic factors that could influence de novo emergence and the details of its underlying mechanism are still unclear.

Answering these questions will help us understand why some parts of the genome transition from functionless to functional while others don't. Knowing how likely such a transition is for different genomic

sequences could also have implications outside of the strict evolutionary context by allowing us to better predict the outcome from future human interventions in the genome such as gene therapies and transgenesis.

We developed a comprehensive methodology combining extensive homology searches, protein evolution simulations and machine learning to, for the first time, reliably identify de novo genes across multiple genomes from the 2 densely sampled yeast genera, *Lachancea* and *Saccharomyces sensu stricto*. We found that, although the rate of de novo emergence is steady within each genus, as previously reported in primates [5], it differs by a factor of 10 between the 2 genera. Young de novo genes are significantly more GC-rich than old, conserved genes and evidence suggests that de novo genes have originated from non-coding regions with significantly higher GC content than the genome average. De novo genes were predominantly found divergently oriented relative to their neighbors, suggesting that their emergence could have been driven by divergent transcription from bidirectional promoters. In *Saccharomyces*, where data are available, de novo genes were also found significantly associated to recombination hotspots. Our results along with the knowledge that in *Saccharomyces,* Double Strand Breaks (DSBs) target conserved promoter regions, lead us to propose that, in yeasts, the combination of 1) high GC% sequences around bidirectional promoters, a result of GC-biased gene conversion following DSBs, and 2) constant divergent transcription of these same regions, provide conditions that favor de novo gene emergence by generating transcripts with lower probability for AT-rich stop codons. High GC% in these transcripts could also lead to (or simply indicate) higher transcriptional or translational efficiency [24-26] which would further increase their chances of emergence.

References

1. Knowles, D. G. & McLysaght, A. Recent de novo origin of human protein-coding genes. Genome Res. (2009). doi:10.1101/gr.095026.109
2. Li, C.-Y. et al. A Human-Specific De Novo Protein-Coding Gene Associated with Human Brain Functions. PLOS Comput Biol 6, e1000734 (2010).
3. Guerzoni, D. & McLysaght, A. De Novo Origins of Human Genes. PLoS Genet 7, e1002381 (2011).
4. Wu, D.-D., Irwin, D. M. & Zhang, Y.-P. De Novo Origin of Human Protein-Coding Genes. PLoS Genet 7, e1002379 (2011).
5. Guerzoni, D. & McLysaght, A. De novo genes arise at a slow but steady rate along the primate lineage and have been subject to incomplete lineage sorting. Genome Biol. Evol. evw074 (2016). doi:10.1093/gbe/evw074
6. Murphy, D. N. & McLysaght, A. De Novo Origin of Protein-Coding Genes in Murine Rodents. PLOS ONE 7, e48650 (2012).
7. Heinen, T. J. A. J., Staubach, F., Häming, D. & Tautz, D. Emergence of a New Gene from an Intergenic Region. Curr. Biol. 19, 1527–1531 (2009).
8. Cai, J., Zhao, R., Jiang, H. & Wang, W. De Novo Origination of a New Protein-Coding Gene in Saccharomyces cerevisiae. Genetics 179, 487–496 (2008).
9. Li, D. et al. A de novo originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand. Cell Res. 20, 408–420 (2010).
10. Wilson, B. A. & Masel, J. Putatively noncoding transcripts show extensive association with ribosomes. Genome Biol. Evol. 3, 1245–1252 (2011).
11. Xiao, W. et al. A Rice Gene of De Novo Origin Negatively Regulates Pathogen-Induced Defense Response. PLOS ONE 4, e4603 (2009).
12. McLysaght, A. & Guerzoni, D. New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. Phil Trans R Soc B 370, 20140332 (2015).
13. Schlötterer, C. Genes from scratch – the evolutionary fate of de novo genes. Trends Genet. 31, 215–219 (2015).
14. Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. & Bosch, T. C. G. More than just orphans: are taxonomically-restricted genes important in evolution? Trends Genet. 25, 404–413 (2009).
15. Li, D., Yan, Z., Lu, L., Jiang, H. & Wang, W. Pleiotropy of the de novo-originated gene MDF1. Sci. Rep. 4, 7280 (2014).
16. Levine, M. T., Jones, C. D., Kern, A. D., Lindfors, H. A. & Begun, D. J. Novel genes derived from noncoding DNA in Drosophila melanogaster are frequently X-linked and exhibit testis-biased expression. Proc. Natl. Acad. Sci. 103, 9935–9939 (2006).
17. Zhao, L., Saelao, P., Jones, C. D. & Begun, D. J. Origin and Spread of de Novo Genes in Drosophila melanogaster Populations. Science 343, 769–772 (2014).
18. Begun, D. J., Lindfors, H. A., Kern, A. D. & Jones, C. D. Evidence for de Novo Evolution of Testis-Expressed Genes in the Drosophila yakuba/Drosophila erecta Clade. Genetics 176, 1131–1137 (2007).
19. Begun, D. J., Lindfors, H. A., Thompson, M. E. & Holloway, A. K. Recently Evolved Genes Identified From Drosophila yakuba and D. erecta Accessory Gland Expressed Sequence Tags. Genetics 172, 1675–1681 (2006).
20. Lin, B. et al. PART-1: a novel human prostate-specific, androgen-regulated gene that maps to chromosome 5q12. Cancer Res. 60, 858–863 (2000).

21. Suenaga, Y. et al. NCYM , a Cis -Antisense Gene of MYCN , Encodes a De Novo Evolved Protein That Inhibits GSK3beta Resulting in the Stabilization of MYCN in Human Neuroblastomas. PLOS Genet 10, e1003996 (2014).
22. Samusik, N., Krukovskaya, L., Meln, I., Shilov, E. & Kozlov, A. P. PBOV1 Is a Human De Novo Gene with Tumor-Specific Expression That Is Associated with a Positive Clinical Outcome of Cancer. PLOS ONE 8, e56162 (2013).
23. McLysaght, A. & Hurst, L. D. Open questions in the study of de novo genes: what, how and why. Nat. Rev. Genet. 17, 567–578 (2016).
24. Gu, W., Zhou, T. & Wilke, C. O. A Universal Trend of Reduced mRNA Stability near the Translation-Initiation Site in Prokaryotes and Eukaryotes. *PLOS Computational Biology* 6, e1000664 (2010).
25. Kudla, G., Lipinski, L., Caffin, F., Helwak, A. & Zylicz, M. High Guanine and Cytosine Content Increases mRNA Levels in Mammalian Cells. *PLOS Biology* 4, e180 (2006).
26. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in Escherichia coli. *Science* 324, 255–258 (2009).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-243

De novo evolved genes are essential for spermatogenesis in D. melanogaster

Jonathan Schmitz [1], Anna Gubala [2], Michael Kearns [2], Tery Vinh [2], Purva Rumde [2], Haylie Butler [2], Emily Rivard [2], Erich Bornberg-Bauer [1], Geoffrey Findlay [2,*]

[1]Institute for Evolution and Biodiversity, University of Muenster, Muenster, Germany, [2]Department of Biology, College of the Holy Cross, Worcester, United States

Abstract: Over the past decade, *de novo* genes that evolved from non-coding DNA sequences have been identified in organisms ranging from plants to humans. Many of these genes are expressed specifically in male reproductive organs, but their functions remain largely unexplored. Here, we describe a bioinformatic screen for *de novo* and putative *de novo* genes expressed in the testes of *Drosophila melanogaster* and use testis-specific RNA interference to investigate each gene's reproductive function in males. These experiments have identified at least three genes that have become essential for male fertility, and several others with more subtle effects. Individual RNAi knockdown and CRISPR knockout of the essential genes leads to a variety of phenotypes. For example, males depleted for one gene fail to produce any sperm, while males knocked out for another show subtle spermiogenesis defects that result in sperm that are unable to localize properly in the female reproductive tract. Molecular evolutionary analyses of these *de novo* genes reveal contrasting patterns of selection. One gene has evolved under positive selection, gene duplication, and genome rearrangement, while others have been conserved through purifying selection since their *de novo* origins. Finally, we examine the relationships between the ages of these newly evolved genes, their genomic locations, their specificity of expression, and their essentiality in male reproduction. Our results suggest that *de novo* and putative *de novo* genes may be important modifiers of male reproductive phenotypes in the face of sexual selection pressures and highlight the importance of conducting genetic tests to interrogate *de novo* gene function.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-244

Young splice sites in primates: positive selection and splicing changes

Stepan Denisov [1,*], Svetlana Iarovenko [2], Mikhail Gelfand [1]

[1]Research and Training Center on Bioinformatics, Institute for Information Transmission Problems, [2]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russian Federation

Abstract: Exon-intron boundaries are marked by splice sites (SSs) – special sequences recognized by spliceosome. Both donor splice site (DSS, at 5' end of intron) and acceptor splice site (ASS, at 3' end of intron) contain key dinucleotide (GT in DSS and AG in ASS) which is surrounded by more variable context. We considered events of new splice site formation (where new key dinucleotide appears) on human lineage after divergence from rhesus macaque. We found 1500 of ASSs and 1244 DSSs created *de novo* on human lineage. About one fifth of these SSs lie within protein coding regions, the remaining ones are located within non-coding RNAs and UTRs. Formation of key dinucleotides is strongly associated with positive selection of wider SS context to increase adherence of SS to spliceosome. Appearance of young SSs changes splicing pattern of corresponding gene. We classified observed changes in splicing and found that young splice sites frequently create new exons from intronic sequences, extend or shorten existing exons, and create new introns within exons. Young splice sites also change first exon of the gene. In cases of new exon creation and extension of the existing exons newly acquired coding sequences tend to be multiple of three by length and seem to be under positive selection on amino acids.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI11

A mouse-specific de novo gene has important functions in developing brain by affecting transcriptional networks

Chen Xie [1,*], Cemalettin Bekpen [1], Sven Künzel [1], Jorge Ruiz-Orera [2], Diethard Tautz [1]

[1]Max Planck Institute for Evolutionary Biology, Plön, Germany, [2]Hospital del Mar Research Institute, Barcelona, Spain

Abstract: For decades, people believed that the evolutionary innovation is from the changes in existing genic and regulatory sequences and the new genes originating through duplications from ancestral genes. In recent years, *de novo* genes, the molecular entities which originated recently in evolutionary history of a given lineage from non-genic ancestral sequences, were found in many species, but their functions and how they affect global gene networks have not been extensively addressed in previous studies. We computationally identified and annotated 483 mouse-specific *de novo* protein-coding genes based on complete genome sequences, whole genome sequencing data, RNA-Seq data, proteomics and ribosome profiling data. We generated knockout mouse lines of three *de novo* genes using homologous recombination and CRISPR/Cas methods. One of them expresses relatively high in postnatal brains. Its open reading frame originated in the past 0.5 million years, and contains 501 nucleotides. We found more than 1,000 differentially expressed genes based on the high resolution RNA-Seq data from the postnatal 0.5 day heads of knockout and wildtype mice. They are statistically significantly enriched in extracellular matrix and regulation of muscle contraction pathways. In addition, the body weight of heterozygous knockout mouse is statistically significantly higher than that of the homozygous knockout mouse. Our results show there are many candidates for clear *de novo* genes in the mouse genome, and one *de novo* gene has important functions in developing brain by affecting transcriptional networks.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI9

Proto-genes, fitness, and de novo gene birth

Anne-Ruxandra Carvunis*

Abstract: Protein-coding genes are thought to emerge *de novo* when non-genic sequences become transcribed, acquire open reading frames, and the corresponding non-genic transcripts access the translation machinery. However, biochemistry predicts that the polypeptides resulting from such translation events should predominantly encode insignificant polypeptides rather than proteins with specific biological roles. It is hard to imagine how such polypeptides could drive species-specific adaptations.

To resolve this conundrum, we have formalized a model according to which the translation of non-genic transcripts does not systematically produce *de novo* genes with adaptive impact, but rather yields transitory "proto-genes". These proto-genes would provide the organism with adaptive potential by exposing genetic variations that are usually hidden in non-genic sequences. The majority would likely return to a non-genic state but a subset may evolve into *de novo* genes, for instance if their expression is beneficial to the organism. In support for this model, widespread translation of non-genic transcripts has recently been documented across numerous species. Whether these translation events really do carry adaptive potential has not yet been demonstrated.

Here, we experimentally assessed the fitness impact of proto-gene expression in the yeast *Saccharomyces cerevisiae*. We systematically overexpressed hundreds of proto-genes and measured the fitness of each resulting strain relative to wild type over multiple environmental conditions. In line with our model, we found that overexpression of proto-genes is mostly neutral and rarely deleterious. Overexpression of proto-genes provided the organism with a beneficial growth advantage significantly more often than the overexpression of canonical genes did across all conditions tested. Strikingly, almost 40% of proto-genes tested showed a beneficial overexpression phenotype when the environment was drastically sub-optimal for the wild type. Our results strongly support the proto-gene model for *de novo* gene birth.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-239

Not so restricted as thought before: an oxygen-binding protein, hemerythrin, in Metazoa

Elisa Costa-Paiva [1,*], Carlos Schrago [1], Kenneth Halanych [2]

[1]Genetics, UFRJ, RIO DE JANEIRO, Brazil, [2]Auburn University, Auburn, United States

Abstract: Recent studies demonstrated that oxygen-binding proteins diversity was underestimated, and despite extensive research on hemoglobins and hemocyanins, little is known about hemerythrins (Hr). Hr homologues are present in the three domains of life, however, in animals, Hr records are restricted for marine invertebrates, belonging to Annelida, Brachiopoda, Priapulida, Bryozoa, and a single species of both Cnidaria and Arthropoda. Given this observed Hr distribution, whether all metazoan Hrs share a common origin is debated. To examine Hr diversity in animals and to further understand Hr evolutionary history, we employed *in silico* approaches to survey metazoan transcriptomes. Sequences of 58 putative Hr genes were identified from 52 species in 10 phyla, being new six records: Mollusca, Echinodermata, Hemichordata, Platyhelminthes, Nemertea, and Phoronida. The presence and expression of Hr genes in referred phyla contradict previous assumptions that Hr genes were absent in the ancestor to the deuterostomes and conserved only in a few protostomes after the protostome-deuterostome split. The topology of the Hr gene tree did not mirror the phylogeny of metazoans as presently understood, which suggests that Hr evolutionary history is hindered by a complex history of gene losses and duplications, paralogous replacements and lateral gene transfer events. Moreover, Bayesian inference revealed two major clades, a clade of metazoan Hrs and a clade of exclusive annelid Hrs, which consists of molecules with a five codon deletion between last alpha-helices. Although our analysis was limited to expressed-coding regions, our findings demonstrate a much greater diversity of Hrs than previously reported.

Expanded summary*: The evolution of metazoans was constrained by the oxygen requirements of tissues. Therefore natural selection has presumably favored proteins that can reversibly bind and transport oxygen (Terwiliger, 1998; Schmidt-Rhaesa 2007). In metazoans, four families of oxygen-binding proteins are known, usually divided into two main groups: proteins that use iron to bind oxygen, including hemoglobins and hemerythrins, and two non-homologous families of hemocyanins that use copper (Burmester 2002). Although these molecules can reversibly bind oxygen, their binding affinities and evolutionary origins differ and recent studies demonstrated that the blood pigments diversity were underestimated in animals (Martin-Duran et al., 2013; Koch et al., 2016). While hemoglobins and hemocyanins have been extensively investigated, knowledge on the evolutionary history of hemerythrins is limited.

Hemerythrin homologues (Hrs) have an ample biological distribution and are present in the three domains of life (Alvarez-Carreño et al., 2016). However, in animals, Hrs records is restricted for marine invertebrates, belonging to Annelida (which include sipunculids), Brachiopoda, Priapulida, Bryozoa, and a single species of both Cnidaria (Nematostella vectensis) and Arthropoda (Calanus finmarchicus) (Vanin et al. 2006; Bailly et al. 2008; Martín-Durán et al. 2013). Bailly et al (2008) suggested that the Hr gene was lost in the ancestor to the deuterostomes and conserved only in a few protostomes after the protostome-deuterostome split. Given this observed Hr distribution in animals, whether all metazoan Hrs share a common origin is debated (Martín-Durán et al. 2013).

Due to the lack of information about Hrs in general, the occurrence and diversity of these molecules may be higher than currently recognized. Thus, to examine a representative sample of metazoan taxa for Hrs and to further understand how different forms of Hrs are evolutionarily related to each other, we employed approaches to survey Hrs from a diverse array of metazoan transcriptomes in silico. We identified sequences of 58 putative Hr genes were identified from 52 animal species in 10 phyla, being new six records: Mollusca, Echinodermata, Hemichordata, Platyhelminthes, Nemertea, and Phoronida. The presence and expression of Hr genes in referred phyla contradict the previous assumption that Hr genes were absent in the ancestor to the deuterostomes and conserved only in a few protostomes

after the protostome-deuterostome split.
Alvarez-Carreño C, Becerra A, Lazcano A. 2016. Molecular evolution of the oxygen-binding hemerythrin domain. PLoS ONE 11(6): e0157904.
Bailly X, Vanin S, Chabasse C, Mizuguchi K, Vinogradov SN. 2008. A phylogenomic profile of hemerythrins, the nonheme diiron binding respiratory proteins.
BMC Evol Biol. 8:244.
Burmester T. 2002. Origin and evolution of arthropod hemocyanins and related proteins. J Comp Physiol B. 172(2):95–107.
Koch J, Lüdemann J, Spies R, Last M, Amemiya C T, Burmester T. 2016. Unusual diversity of myoglobins genes in the lungfish. Mol Biol Evol.
doi:10.1093/molbev/msw159.
Martín-Durán JM, De Mendoza A, Sebé-Pedrós A, Ruiz-Trillo I, Hejnol A. 2013. A broad genomic survey reveals multiple origins and frequent losses in the
evolution of respiratory hemerythrins and hemocyanins. Genome Biol Evol. 5:1435–1442.
Schmidt-Rhaesa A. 2007. The Evolution of Organs Systems. Oxford University Press. 385 pp.
Terwilliger NB. 1998. Functional adaptations of oxygen-transport proteins. J Exp Biol. 201:1085-1098.
Vanin S, Negrisolo E, Bailly X, Bubacco L, Beltramini M, Salvato B. 2006. Molecular evolution and phylogeny of sipunculan hemerythrins. J Mol Evol.
62:32-41.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-240

GenTree as an integrated resource database for evolutionary new genes in humans

Yi shao [1], Chunyan chen [1,*], Hao shen [2], Manyuan Long [3], Yong zhang [1]

[1]computational evolutionary genomics, Institute Of Zoology, Chinese Academy Of Sciences, Beijing, [2]Institute of computer and communication, hunan university of technology, hunan, China, [3]Department of Ecology and Evolution, The University of Chicago , Chicago, United States

Abstract: In the recent decade, rapidly accumulating evidence demonstrates that lineage- or species-specific new genes serve as an important player in phenotypic evolution across animal and plant. However, only a dozen cases have been experimentally characterized in humans, although hundreds of primate-specific genes exist in the reference genome. The lack of study is partially due to their relatively poor annotation quality caused by low transcription abundance, narrow expression profile as well as high sequence similarity to their paralogous genes (for duplicated new genes). In order to alleviate the difficulty from these issues in further study of new genes in the genomes of humans, we have developed and improved a user-friendly web interface, GenTree, (http://gentree.ioz.ac.cn). With this web interface, users can browse when and how one gene of interest gets originated. More than that, we integrated transcriptome, proteome and genetic association data into GenTree while taking mapping ambiguity between paralogs into account. Thus, not only can users evaluate whether one gene is more likely coding or non-coding, but also get a hint of its potential functionality. Although it has not been formally published, it has been visited for hundreds of times with dozens of Email inquiries. With these inputs, we continuously update the web interface with the hope of making a more useful resource.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI4

Origination of retrochimera via the LTR-mediated mechanism in metazoans

Shengjun Tan [1], Margarida Cardoso-Moreira [2], Jinbo Wang [1], Yi Shao [1], Chunyan chen [1], Manyuan long [3], Yong Zhang [1,*]

[1]Institute of Zoology, CAS, Bejing, China, [2]Center for Integrative Genomics, University of Lausanne , Lausanne, Switzerland, [3]Department of Ecology and Evolution, The University of Chicago , Chicago , United States

Abstract: Genes can duplicate through an RNA intermediate (*i.e.*, retroposition). This process hasbeen mostly studied in mammals where a non long-terminal-repeat (LTR) element, L1 retrotransposon, captures mRNAs and inserts them back into the genome. However, the elements responsible for retroposition in other animals are less characterized. Therefore, we examined young retrocopies that still retain the features indicative of the underlying mutational mechanism across various animals. In fruitfly, we assembled 15 polymorphic retrocopies and found that all are chimeras of internal retrocopies flanked by LTR retrotransposons. At the fusion points, we observed shared short similar sequences suggesting the involvement of microsimilarity-dependent template switches. By expanding our approach to the reference genomes of mosquito, zebrafish, chicken, and mammals, we found recently fixed retrocopies with a similar chimeric structure. Interestingly, unlike L1-mediated retrocopies, which are usually dead on arrival due to the lack of regulatory sequences, LTR-mediated retrocopies are immediately co-transcribed with their flanking LTR retrotransposons. CRISPR-Cas9 induced GFP tagging experiments further demonstrate that that at least one chimeric retrocopy generates a highly abundant protein. Despite their potential functional importance, similarly to other types of new genes (*e.g., de novo* genes), fixed retrocopies are poorly annotated. Thus, starting with the human genome, we have integrated proteome data and evolutionary inference (*e.g.*, K_A/K_s) to evaluate gene models for recently evolved new genes, and developed a user-friendly database, GenTree (http://gentree.ioz.ac.cn).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-227

Massive duplication of Ribosomal Protein genes was associated with convergent evolution of aerobic fermentation in yeasts

Alison Robson [1,*], Zhenguo Lin [1]

[1]Biology, Saint Louis University, Saint Louis, United States

Abstract: The ribosome is the translation machinery of all cellular organisms. The origin and evolution of ribosomal proteins is of great interests because it provides new insights into understanding about cellular organisms evolved. In this study, to elucidate the evolutionary history of RP genes in fungi, we identified all ribosomal proteins from 18 completely sequenced fungal species and conducted phylogenetic analysis for each RP gene family. We found that most of RP families have experienced gene duplication in many species of two distantly related yeast lineages: the budding yeast and the fission yeast. For instance, 65 of 78 RP genes in the budding yeast *Saccharomyces cerevisiae* and 62 of 79 RP genes in RP genes in the fission yeast *Schizosaccharomyces pombe* have been duplicated. Interestingly, these yeast species with massive RP gene duplication shared a unique physiological trait: aerobic fermentation, the ability to ferment sugar into ethanol even in the presence of oxygen. Aerobic fermentation, which has independently evolved in the two yeast lineages, was suggested to allow these yeasts to out-compete other microorganisms through rapid growth and consumption of sugars. Our analysis suggested that the RP genes in the budding yeasts were likely duplicated by the whole genome duplication (WGD) event. However, since fission yeasts did not undergo a WGD, these RP genes were duplicated individually. The significant correlation between massive duplication of RP genes and aerobic fermentation suggests that the expansion of RP genes might have contributed to the convergent evolution of aerobic fermentation in budding and fission yeasts.

Expanded summary*: Considerable research has been done on the ribosomal proteins in the model yeast *S. cerevisiae* revealing massive retention of the duplications originating from the WGD, subsequent gene conversion and subfunctionalization. The *Schizosaccharomyces* genome project completed in 2011 also revealed similar duplications in a majority of the ribosomal protein genes in all four species of fission yeasts (Rhind et al., 2011). *S. cerevisiae* and its close relatives and the fission yeasts are the only lineages with this level of duplication in their ribosomal proteins. Strikingly, these two lineages share one interesting phenomenon: aerobic fermentation. Aerobic fermentation is preferentially fermenting glucose and other sugars via the fermentation pathway after glycolysis and the repression of the normal respiratory pathway in the presence of oxygen. The evolution of this phenomenon initially appears counterintuitive, as fermentation produces significantly less metabolic energy (ATP) than respiration does per glucose molecule and results in the toxic product, ethanol. However, the evolution of aerobic fermentation is believed to be associated with the emergence of modern fruit (Piškur, Rozpędowska, Polakova, Merico, & Compagno, 2006). It is believed that a fermentative lifestyle was advantageous for these yeasts as they rely solely on simple sugars and have high competition from other yeasts and prokaryotes. Fermentation could allow for faster consumption of sugars and provide a toxic environment for their competitors.

Thus, the switch to an aerobic fermentative lifestyle would also require increased uptake of glucose, increased tolerance to ethanol, and increased growth rate. In this project, we are focusing on the increased growth rate and how it potentially relates to the duplicated ribosomal proteins in aerobic fermentative yeast species. In microbial species, there often exists a trade-off between growth rates and maintenance. When conditions oscillate or are fairly variable, a slower growth rate is evolutionarily favorable (Ying et al., 2015). However, *S. cerevisiae* and other aerobic fermentative yeasts naturally exist in relatively stable high sugar environments (Hittinger, 2013; Williams, Liu, & Fay, 2015). However, since they do aerobic fermentation, their growth rates would theoretically be low as they generate less energy for the sugar they consume. Despite this, aerobic fermentors have similar if not higher growth rates than respiratory yeasts (Hagman, Säll, Compagno, & Piskur, 2013). Several evolutionary events may have contributed to this high growth rate in aerobic fermentative species, including increased uptake of glucose through the expansion of hexose transporters (Lin & Li, 2011b).

Furthermore, the duplication of ribosomal proteins may also play a role in ensuring high growth rates during excess of sugars and increased growth when the ethanol concentration builds up and most of the glucose is consumed. The duplication and retention of RP genes may be useful to ensure a constant dose with one copy compensating for the other in the event of loss or gene instability. A duplicate copy may act as a buffer or backup, which is observed in other duplicated genes as well (Ihmels, Collins, Schuldiner,

Krogan, & Weissman, 2007). Gene conversions could ensure substitutability of the backup copy and the functional differences in the two copies due to the intron sequences could prevent the second copy from being lost. Alternatively, the second copy could be essential during stress conditions or when switching back to respiration as ethanol concentrations rise and the functional discrepancies may favor the appropriate ribosomal assembly and structure for the new needs of the cell. Whether it is for ensuring a constant dose or subfunctionalization, having duplicated RPs may have provided further support for yeasts transitioning into a fermentative lifestyle.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI10

"Fishing" for vertebrate fertilization genes: genomic and proteomic characterization of rapidly evolving threespine stickleback egg proteins

Emily Killingbeck [1,*], Damien Wilburn [1], Gennifer Merrihew [1], Michael MacCoss [1], Catherine Peichel [2], Willie Swanson [1]

[1]Genome Sciences, University of Washington, Seattle, United States, [2]Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

Abstract: Sperm-egg compatibility is essential to the evolutionary success of any sexually reproducing organism, yet the proteins that mediate gamete interactions often evolve at extraordinary rates. In threespine stickleback fish (*Gasterosteus aculeatus*), reproductive isolation is common in many recently derived populations throughout the Northern Hemisphere, but the precise biochemical mechanisms driving this isolation are unknown. Stickleback are classic models of molecular adaptation and speciation, and while rapidly evolving reproductive proteins are probable candidates underlying this reproductive isolation, they remain unexplored in this model evolutionary system. Tandem mass spectrometry was used to characterize the secreted proteomes of stickleback eggs from Lake Union, Washington. High-resolution mass spectra were acquired, with homologs of common vertebrate egg proteins identified. Evolutionary rate analysis (d_N/d_S) of these homologs across fish from superorders within the Teleosts indicates positive selection. In contrast to mammals, the genes encoding the major egg proteins are tandemly duplicated in the stickleback genome. Such duplications provide a substrate for diversification that can drive rapid evolution, and suggest a potential mechanism underlying sexual conflict within stickleback populations and ultimately speciation.

Expanded summary*: Approach

Fertilization is mediated by sperm and egg cell surface proteins. Reproductive proteins are paradoxical: given their fundamentality to propagation, interacting male and female proteins might be expected to be highly conserved to maintain compatibility – instead, however, these genes are generally among the most rapidly evolving in the genomes of most taxa. This juxtaposition of rapid evolution and biological constraint is likely driven by coevolution between interacting reproductive protein partners to maintain compatibility. Further, rapid evolution can be exploited to refine putative reproductive protein candidates as well as to reveal sites of likely functional significance within a protein. The contribution of rapidly evolving fertilization proteins to reproductive isolation and speciation is not well-understood, and we seek to address this question using threespine stickleback fish (*Gasterosteus aculeatus*) as a model system.

Adaptive radiation of ancestral marine stickleback fish into freshwater environments throughout the Northern Hemisphere – often following glacial retreat in the last ~15,000 years – has resulted in phenotypically divergent forms characterized by varying degrees of reproductive isolation. The role of sperm-egg recognition proteins in this reproductive isolation remains an open question, as the consensus in the field has been that behavioral differences in courtship or habitat preferences underlie stickleback population divergence.

In animals, the oocyte is surrounded by a raised egg coat which serves as a barrier to polyspermy. Mate recognition phenotypes are believed to result from the interaction of the protein products of only a few genes on the cell surface of gametes. To determine the molecular mechanisms underlying gamete recognition in stickleback, we are identifying secreted proteins in egg and sperm with shotgun proteomics using tandem mass spectrometry. Preliminary data indicate the proteome of stickleback egg coats is relatively simple, with the most abundant proteins containing zona pellucida (ZP) domains homologous to those found in mammalian egg coats. Evolutionary rate analysis (d_N/d_S) of stickleback ZP homologs across fish from Superorders within the Teleosts indicates positive selection. In many cases, the ZP homologs are tandemly duplicated in the stickleback genome. Tandem repetitive sequence elements provide a substrate for diversification that can drive rapid evolution, suggesting a potential mechanism underlying sexual conflict and ultimately speciation.

Fertilization specificity in stickleback is likely aided by the micropyle, a funnel-like structure that restricts sperm entry through the egg coat. We hypothesize that proteins localized to this region of the egg coat are playing a role in sperm-egg interaction specificity, and seek to isolate micropyles by laser capture microdissection for subsequent proteomic and biochemical characterization, including investigations of glycosylation or other modifications.

Significance to reproductive biology
A comprehensive understanding of reproductive biology cannot be achieved without addressing the molecular mechanisms that mediate fertilization. Up to 20% of cases of human infertility have no known cause. As a vertebrate model system, understanding the process of fertilization in threespine stickleback fish may illuminate important molecular properties that extend to human reproductive biology.

Significance to evolutionary theory
Threespine stickleback fish are a well-established model of molecular adaptation and speciation. In existing studies, pre-zygotic explanations for stickleback reproductive isolation have been limited to behavioral differences in mating behavior or habitat preference. Our preliminary investigations, however, have found evidence of rapid evolution of stickleback egg coat proteins. Our data, which suggest that sperm competition and potentially other molecular forms of pre-zygotic mating isolation may be facilitating stickleback population divergence, challenge the current consensus in the field.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI6

Recurrent gene duplication leads to diverse repertoires of centromeric histones in Drosophila species

Lisa Kursel [1 2,*], Harmit Malik [1 3]

[1]Fred Hutchinson Cancer Research Center, [2]University of Washington, Seattle, [3]HHMI, Chevy Chase, United States

Abstract: Despite their essential role in the process of chromosome segregation in most eukaryotes, centromeric histones show remarkable evolutionary lability. Not only have they been lost in multiple insect lineages, but they have also undergone gene duplication in multiple plant lineages. Based on detailed study of a few model organisms including Drosophila melanogaster, centromeric histone duplication is considered to be rare in animals. Using a detailed phylogenomic study, we find that Cid, the centromeric histone gene, has undergone four independent gene duplications during Drosophila evolution. We find duplicate Cid genes in D. eugracilis (Cid2), in the montium species subgroup (Cid3, Cid4) and in the entire Drosophila subgenus (Cid5). We show that Cid3, Cid4, Cid5 all localize to centromeres in their respective species. Some Cid duplicates are primarily expressed in the male germline. With rare exceptions, Cid duplicates have been strictly retained after birth, suggesting that they perform non-redundant centromeric functions. Indeed, each duplicate encodes a distinct N-terminal tail, which may provide the basis for distinct protein-protein interactions. In addition, we show some Cid duplicates evolve under positive selection whereas others do not. Finally, we have used Cid5-specic antibodies to determine that Cid5 functions post-meiosis and localizes to foci in the nuclei of developing spermatids in D. virilis. These results suggest that Cid5 has acquired a specialized centromeric role in the male germline. Taken together, our results support the hypothesis that Drosophila Cid duplicates have subfunctionalized. Thus, these gene duplications provide an unprecedented opportunity to dissect the multiple roles of centromeric histones.

Expanded summary*: Centromeres are the chromosomal regions that link DNA to the spindle during cell division, thus ensuring faithful segregation of genetic material. The hallmark of many centromeres is the presence of a specialized centromeric H3 variant called CenH3 (Cid in *Drosophila).* CenH3 is essential for chromosome segregation and thus essential for life in most eukaryotes. However, despite its essentiality, *CenH3* and the centromeric DNA sequences to which binds, evolve rapidly. Genetic conflicts provide one potential explanation for the rapid evolution of CenH3. In both animals and plants, the asymmetry of female meiosis provides an opportunity for centromere alleles to act selfishly to favor their own inclusion in the oocyte and subsequent passage into offspring rather than the polar body. In males, however, 'driving centromeres' are thought to result in reduced fertility. This lower fertility is predicted to cause the evolution of genetic suppressors of centromere drive. Under this model, centromeric proteins evolve rapidly in order to mitigate fitness costs associated with centromere drive

Centromere drive and its suppression provide an explanation for the rapid evolution of centromeric DNA and centromeric proteins. However, it invokes the relentless, rapid evolution of essential proteins such as CenH3, whose mutation could be highly deleterious. A simpler way to allow for the rapid evolution of centromeric proteins without compromising their essential function would be via gene duplication. Duplication and specialization of centromeric proteins would allow one paralog to function as a drive suppressor in the male germline, while allowing the other to carry out its canonical centromeric role.

To study the incidence of *CenH3* duplication in a well-studied animal lineage, we took advantage of the recent sequencing of high-quality genomes from multiple *Drosophila* species. Despite there being only one copy of *CenH3* (*Cid1*, previously known as just '*Cid')* in *D. melanogaster*, we were surprised to find that some *Drosophila* species had two or more copies of *CenH3.* Based on BLAST analyses and PCR analyses of non-sequenced genomes, we found that the species of the *montium* subgroup, including *D. kikkawai*, have three *Cid* genes (*Cid1, Cid3 and Cid4),* which were born from a duplication event ~15 million years ago. The species of the *virilis* group, as well as *D. mojavensis* and *D. grimshawi* (*repleta* and *Hawaiian* groups, respectively), have two *Cid* genes (*Cid1* and *Cid5),* which were born from a duplication event ~40 million years ago. These *Cid* duplications have been almost completely preserved in extant species. Despite the fact *Cid* paralogs are divergent from one another at the sequence level, all paralogs have the ability to localize to centromeres when expressed in tissue culture cells. Phylogenetic analyses support our synteny-based conclusions, and reveal recurrent recombination between *Cid1* and *Cid3* in *montium* subgroup species. Our results suggest that recombination results in evolutionary homogenization of the histone fold domain between *Cid1* and *Cid3*, while the N-terminal tails of *Cid1* and *Cid3* appear to be evolving independently, perhaps maintaining divergent functions.

The fact that multiple *Cid* paralogs have been retained for millions of years suggests that these paralogs perform non-redundant roles. We explored this possibility further by examining the expression pattern and selective constrains acting on each *Cid* paralog. Interestingly, we find that some *Cid* paralogs *(Cid3* in the *montium* group and *Cid5* in the *virilis* group) are expressed primarily in the male germline. Furthermore, we find that some *Cid* paralogs evolve under recurrent positive selection at sites predicted to contact centromeric DNA. Lastly, we identified four conserved N-terminal tail motifs that are present in all single copy Cids. We find that these motifs are differentially retained among Cid paralogs, further supporting our subfunctionalization hypothesis.

We propose that in species with a single-copy Cid gene, the same protein must perform multiple functions including mitotic cell division in somatic tissues and drive suppression in the male germline. These functions might require different selective pressures to achieve functional optimality. For example, drive suppression may result in rapid evolution of *Cid* to co-evolve with rapidly evolving centromeric DNA whereas mitotic function might impose purifying selection on *Cid*, minimizing changes in amino acid sequence. Therefore, it could be advantageous to have two copies of *Cid* such that each encodes a separate function. Our results suggest the intriguing possibility that *CenH3* duplications may allow *Drosophila* species to better achieve functional optimality of multiple centromeric functions than species encoding a single *CenH3* gene.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI1

Investigating young de novo genes in Drosophila melanogaster.

David Begun*, Li Zhao [1]

[1]Rockefeller University, New York, United States

Abstract: De novo genes, which result from the evolutionary transition of DNA sequences from non-genic to genic, are among the most unusual of all genomic oddities, as their derived biological functions are likely completely unrelated to their ancestral states. Much of our recent work has sought to illuminate the earliest stages in the evolution of such genes through the investigation of young testis-expressed de novo genes in Drosophila melanogaster. In that work we discovered a substantial number of testis-biased genes that had fixed in the *D. melanogaster* lineage since its split from the *D. melanogaster-D. simulans* common ancestor or are segregating in *D. melanogaster* populations. Here I will present some of our recent work extending these early studies to additional tissues and lineages to gain a broader view of the abundance, potential biological properties, and population genetics of de novo genes. Preliminary results suggest that the testis is indeed highly unusual compared to other tissues regarding the proportion of expressed genes having de novo origin.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-247

The molecular underpinnings of adaptation in a novel defensive gene in tomato (Solanum lycopersicum)

Allan Castillo [1,*], Mark D. Rausher [1]

[1]Biology, Duke University, Durham, United States

Abstract: A major goal of molecular evolution is to determine how evolutionary novelty arises following gene duplication. The amino-acid biosynthetic enzyme threonine deaminase (TD) duplicated early in the evolution of the Solanaceae plant family, giving rise to a copy with a novel defensive function (TD2). Unlike the ancestor-like copy (TD1), TD2 evolved to thrive in the harsh alkaline gut environment of lepidopteran herbivores and to deplete the herbivores' of the essential amino acid threonine. Studies of the TD copies in tomato (Solanum lycopersicum) have shown that three critical ion pairs—interacting pairs of oppositely charged amino acid residues—exist in the defensive copy of TD2 but not in TD1, and are thought to play a role in increasing thermostability. We hypothesize these ion pairs stabilize TD2 in environments with high pH levels, high heat, and resist digestion by proteases to become an effective antinutritive enzyme. Here we investigate the effects of each critical ion pair by disrupting the ion pairs with site-directed mutagenesis, converting one of the residues to the most recent ancestral state, and performing enzymatic assays in-vitro feeding assays with the Solanaceae generalist herbivore Manduca sexta. While other amino-acids may contribute the TD2's stability, high pH tolerance, and protease resistance, these may convey a large effect on giving TD2 its defensive capabilities.

Expanded summary*: Models of functional innovation following gene duplication are often invoked to explain the evolution of novel protein function. Though gene duplication is often hypothesized to provide the basis for evolutionary innovation, we have a very limited understanding of the evolutionary process and mechanisms that underlie adaptive evolution following gene duplication. This investigation of adaptive innovation follows the duplication of the gene threonine deaminase (TD) in tomato (*Solanum lycopersicum*). The TD1 copy retains the ancestral function in catalyzing threonine as part of the isoleucine biosynthesis pathway. A duplicate copy, TD2, evolved into a defensive enzyme against lepidopteran herbivores by functioning as an antinutritive in the gut. Relative to TD1, the defensive TD2 enzyme is much more stable, tolerant of higher pH levels, and resistant to protease digestion. These adaptations allow TD2 to survive and thrive in the lepidopteran gut. Previous research showed that TD2 contains three extra critical ion pairs, linkage structures that stabilize proteins and may provide thermostability to the TD2 enzyme. I can test this hypothesis using both phylogenetic and molecular approaches, including direct disruption of the ion-pairs. My analysis of the TD gene tree in Solanaceae shows positive selection along the TD2 branch leading to tomato, whereas TD1 has experienced purifying selection since the duplication event, consistent with the neofunctionalization model. It is possible that the critical ion pairs were selected in TD2 to provide protein stability and high pH resistance needed to survive in the lepidopteran herbivore gut and evolve its defensive capabilities.

Significance: Despite the many models of evolutionary novelty (neofunctionalization, escape from adaptive conflict, subfunctionalization, etc.), few studies provide empirical evidence of the functional and molecular processes leading to adaptive innovations. Specifically, contingency, epistasis, and the effects of different amino acid substitutions may play significant roles in the evolutionary trajectory of novel genes. This study assesses the role that critical ion pairs play in the evolution of a novel gene, which has scarcely been researched, to determine whether stability and high pH tolerance was necessary to evolve into its new function. Investigating how critical ion pairs stabilize TD2 will reveal the interplay of biochemistry and molecular evolution for understanding protein evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-246

Identification of an alternative survival mechanism in fly-vectored yeast

Steven Dittmer [1], Kate McKee [1,*], Kelly M. Thomasson [1], Stephen R. Proulx [1], Lena Dominguez-Meneses [1]

[1]Ecology, Evolution and Marine Biology, University of California , Santa Barbara, United States

Abstract: Organisms adapt to their environments in order to best survive and propagate. In the case of the yeast, *Saccharomyces cerevisiae*, a key aspect of survival is the process of sporulation, during which the vegetative cell ceases mitotic division and forms a protective outer-coating. Previously, spore formation has been described as a critical adaptive mechanism for survival of an insect's digestive tract when these eukaryotic microbes are dispersed by insect gut-vectoring. However, there may be alternative strategies for survival of this digestion process. One such alternative strategy is observed in several strains of *S. cerevisiae* in which the vegetative cells excrete a protective, agglutinating substance that may render obviate spore formation. We performed a series of biochemical analyses on this substance to determine its molecular composition. Through comparative protein analysis and spectroscopy we were able to better understand the composition of the substance and postulate mechanisms that allow for protection and survival. Results of this analysis provide a more complete picture of the diversity of phenotypes within the *Saccharomyces* complex and help us to understand how yeast have responded to selection pressure imposed by insect gut-vectoring.

Statement: I am currently a fourth-year biological sciences student at the University of California, Santa Barbara. For the past year, I have been participating in evolutionary biology research in Dr. Stephen Proulx's lab on campus. My independent research involves studying the regional phenotypic differences in fly-vectored Saccaromyces cerevisae as well as the survivorship of vegetative and sporulated yeast cells processed through the Drosophila gut. I am very interested in the incorporation of genetics into evolutionary studies. The depth of my knowledge in this subject that I have gained from the research I have taken part in has greatly surpassed anything I have learned in a classroom. The opportunity to attend the 2017 SMBE meeting would further expand my interest in the field of genetics and molecular biology and serve as a capstone to my undergraduate experience. I am looking forward to attending talks given by the brightest scientists in our field discussing novel topics and answering questions I have never even thought to ask.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-245

Most yeast de novo genes evolve from transcript isoforms of ancient genes

Tzu-Chiao Lu [1 2 3,*], Wen-Chang Lin [2 3], Jun-Yi Leu [1 2]

[1]Institute of Molecular Biology, Academia Sinica, [2]Graduate Institute of Life Sciences, National Defense Medical Center, [3]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

Abstract: *De novo* genes, novel genes arising from random DNA sequences, have been suggested to be widespread on the genome among different organisms. However, our knowledge about the origination and evolution of *de novo* genes is still limited. To systematically understand the general features of *de novo* genes, we established a robust pipeline to analyze more than 20,000 transcript-supported coding sequences (CDSs) from *Saccharomyces cerevisiae*. Our analysis pipeline combined phylogeny, synteny, and sequence alignment information to identify possible orthologs across 20 *Saccharomycetaceae* yeasts and discovered 560 *S. cerevisiae*-specific *de novo* genes and 1731 *S. sensu stricto*-specific *de novo* genes. We further combined information of CDS positions and transcript structures to show that more than 75% of *de novo* genes arose from transcript isoforms of ancient genes, especially in the upstream and internal regions of ancient genes. Fourteen identified *de novo* genes with high transcript levels were chosen to verify their protein expression. Ten of them, including 8 transcript-isoform-associated CDSs, showed translation signals and 5 proteins exhibited specific cytosolic localizations. Our results suggest that *de novo* genes frequently arise in the *S. sensu stricto* complex and have the potential to be quickly integrated into the ancient cellular pathways.

Expanded summary*: Recent comparative genomic studies have revealed that most eukaryotic genomes contain sets of genes that are not found in other related species. These “taxonomically-restricted genes” or “novel genes” have been suggested to play important roles when organisms adapt to different habitats. Taxonomically-restricted genes arise through various mechanisms, including gene duplication, retrotransposition, exon shuffling, horizontal gene transfer, gene fission/fusion and de novo origination. Although most mechanisms for novel-gene creation have been extensively studied, our understanding of *de novo* formation remains limited. In this poster, we discuss the possible mechanism for *de novo* gene formation in *S. sensu stricto* yeasts. Combing the evolution history and transcription profile of *de novo* genes, we show that most yeast *de novo* genes evolve from transcript isoforms of ancient genes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-242

Genomic conflict drives the evolution of gene copy number and reproductive barriers between species

Emily Kopania [1,*], Erica Larson [1], Jeffrey Good [1]

[1]University of Montana, Missoula, United States

Abstract: Genomic conflict, in which selfish genes act to increase transmission to offspring at a cost to the organism, plays an important role in genome evolution. In mice, genomic conflict is thought to have driven gene amplification on the sex chromosomes, resulting in large copy number variants (CNVs). Sex-linked CNV amplification primarily occurs in genes expressed during postmeiotic spermatogenesis when the sex chromosomes are repressed. Relative postmeiotic CNV expression levels between the X and Y chromosomes have been linked to sex-ratio distortion, thereby antagonistically linking CNV evolution with X-Y conflict. While autosomes do not undergo postmeiotic repression, many autosomal genes appear to be coregulated with the sex chromosomes during spermatogenesis, but the extent to which genomic conflict drives autosomal CNV evolution is largely unexplored. To address this, we have combined comparative genomic analyses of CNV sequence evolution with novel cell-specific postmeiotic transcriptome profiles across several species of mice. We are using these data to identify networks of postmeiotic gene expression and to test the hypothesis that autosomal and sex-linked CNVs are coevolving due to genomic conflict. Our studies provide novel insights into the functional importance of genomic conflict on genome evolution and the evolution of interspecies reproductive barriers.

Expanded summary*: Examples of genomic conflict are abundant, but its significance in genome evolution remains largely unknown[1]. In mice, genomic conflict is thought to have driven Y-chromosome evolution[2,3,4]. Compared to many other mammals, the mouse Y-chromosome is long, largely composed of euchromatin, and enriched for ampliconic genes[2]. Antagonistic coevolution between mouse sex chromosome genes has been shown to play an important role in the evolution of copy number variants (CNVs). Increased copy number results in increased expression levels despite repression during postmeiotic spermatogenesis[3,4]. While whole chromosome repression does not occur on the autosomes, several autosomal genes are coregulated with the sex chromosomes during spermatogenesis[3,5]. These autosomal regions may be evolving due to genomic conflict with the sex chromosomes during spermatogenesis. We are using leading bioinformatics and genomics approaches to gain new insight into the importance of genomic conflict in genome evolution and copy number variation.

This project is also investigating the role of genomic conflict in causing reproductive isolation. Antagonistic coevolution appears to be driving copy number divergence between species. Therefore, hybrid male mice with CNVs from different species have a copy number imbalance, resulting in malformed sperm and sterility[3,5]. This model predicts that coevolving CNVs will have similar disrupted expression patterns during spermatogenesis in sterile hybrids. We are testing this using innovative single-cell gene expression analyses. This project provides novel understanding of the importance of genomic conflict in genome evolution and the functional effects of this genomic divergence on reproductive barriers between species. 1: Rice, W. R. (2013). *Annu Rev Ecol Evol Syst.* **44**(1):217-237. 2.Soh, Y. Q., et al. (2014). *Cell* **159**(4):800-813. 3:Cocquet, J., et al. (2012). *PLoS Genet* **8**(9):e1002900. 4:Ellis, P. J., et al. (2011). *Hum Mol Genet* **20**(15):3010-3021. 5:Larson, E.L., et al. (2016). *Mol Biol Evol. In press.*

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-252

Genome defense innovation via pervasive APOBEC3 retroduplications in primates

Richard McLaughlin [1,*], Harmit Malik [1], Michael Emerman [2]

[1]Division of Basic Sciences, [2]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, United States

Abstract: The replication of mobile genetic elements imposes fitness costs on hosts. Consequently, host genomes evolved an array of restriction factors to block the replication and deleterious effects of these elements. However, the composition of a host's pathogen repertoire changes over evolutionary time, as there is always an advantage to be gained by pathogens in evading host restriction. The APOBEC3 (A3) family of cytidine deaminase genes provides restriction of infectious and endogenous pathogens. The dynamic expansion and contraction of this gene family likely represents host adaptation to keep apace of the dynamic evolution of pathogens. For example, we find that there has been a dramatic expansion of the A3 gene family in New World monkeys, with up to nine additional copies of A3G in some species; however, this expansion occurred via retroposition, in which an endogenous retroelement duplicates a host gene via its transcribed RNA. We find these duplications to be pervasive in New World monkeys and other primates. Further, we see that some A3 retrogenes are ancient and some have been born very recently. Finally, we show that some of these gene copies encode important genome defense functions including LINE-1 and retrovirus restriction. We suggest that retroposition of host defense factors represents an unappreciated mechanism of rapid adaptation to block pathogens. In this way, the host genome co-opts retroelement activity to block retroelement replication.

Expanded summary*: I study the impact of selfish elements on human biology using diverse computational and wet lab techniques. In my postdoctoral training, I developed a research program to study the evolutionary conflicts between primate genomes and their endogenous retroelements. Co-mentored by virologist Michael Emerman, I showed that a potent LINE-1 restriction factor, *APOBEC3A*, adapted to restrict viruses in a manner that preserves LINE-1 restriction. This work exemplifies the tradeoffs and flexibility inherent in the adaptation of host defense genes that must restrict multiple pathogens. I also demonstrated how host genomes utilize retroelement sequence and activity to generate novel mechanisms of genome defense. In collaboration with evolutionary biologist Holly Wichman, I showed how *L1TD1*, a human pluripotency gene, was born from the domestication of a LINE-1 retroelement and may function as a restriction factor against the very LINE-1 retroelements from which it was domesticated. In the presented abstract, I show that the retroposition activity of LINE-1 retroelements led to the generation of many new *APOBEC3G* genes in New World monkeys. Some of these genes were born at least 60 million years ago, at least in the common ancestor of simian primates, and some of these genes are so recently born that I see variation in the presence of these genes within populations of certain species. In addition to this interesting phenomenon, a subset of these newly formed genes are expressed in germline and somatic tissues and potently block LINE-1 and retroviruses. We suggest that this process may be a mechanism to rapidly generate novelty in innate immune defense against both endogenous and infectious pathogens. Finally, we propose that the gene duplicating activity of retroelements like LINE-1 may produce a feedback loop in which LINE-1 activity increases the repertoire of host anti-LINE-1 restriction factors. This adds another dimension to the demonstration that hosts take advantage of the massive portion of their genomes that derive from transposable elements as a source of raw materials (sequences and enzymatic activity) to innovate host biology.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

OT-MI5

How and why are there so many introns in genomes? Mechanism by which transposable elements generate introns on massive scales

Jason Huff [1,*], Daniel Zilberman [1], Scott Roy [2]

[1]University of California, Berkeley, Berkeley, CA, [2]San Francisco State University, San Francisco, CA, United States

Abstract: The discovery of introns four decades ago was one of the most unexpected findings in molecular biology. Introns are sequences interrupting genes that must be removed as part of messenger RNA production. Genome sequencing projects have shown that most eukaryotic genes contain at least one intron, and frequently many. Comparison of these genomes reveals a history of long evolutionary periods during which few introns were gained, punctuated by episodes of rapid, extensive gain. However, although several detailed mechanisms for such episodic intron generation had been proposed, none had been empirically supported on a genomic scale.

We built upon our previous characterization of a novel DNA methylation system in eukaryotes[1], related to nucleosome positioning within chromatin, to elucidate the mechanism by which introns are generated. We found that short, non-autonomous DNA transposable elements independently generated hundreds to thousands of introns in the prasinophyte *Micromonas pusilla* and the pelagophyte *Aureococcus anophagefferens*[2]. Each transposable element carries one splice site. The other splice site is co-opted from the gene sequence that is duplicated upon element insertion, allowing perfect splicing out of the RNA.

The two independent examples of proliferating elements illustrate a general DNA transposable element mechanism that can plausibly account for episodes of rapid, extensive intron gain during eukaryotic evolution. Numerous additional implications for genome evolution will be discussed. This mechanism potentially solves the long-standing mystery of *how* and begins to address *why* there are so many introns.

[1]Huff JT, Zilberman D. ***Cell*** 2014; 156(6): 1286-97.
[2]Huff JT, Zilberman D, Roy SW. ***Nature*** 2016; 538(7626): 533-6.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Molecular innovation

POA-249

Evolution of a new gene: the problem of self-tolerance versus autoimmunity

Cemalettin Bekpen*, Chen Xie, Diethard Tautz

Abstract: In vertebrates, the adaptive Immune system has an extraordinary potential for generating receptors that sense and neutralize any foreign antigens entering the body. Efficient recognition of the foreign antigens depends on the regulation in Thymus tissue where T cells are selected first positively and negatively. Promiscuous gene expression of Tissue specific antigens (TSA), which is required for negative selection, is introduced to T cells by mature mTEC cells mostly through the control of *Aire* gene function in Thymus. Any failure within the function of the *Aire* gene results in the loss of sense and non-sense separation and thereby autoimmunity due to improper representation of TSA in Thymus.
De novo evolved genes usually bring novel expression pattern with newly evolved genetic content to specific tissues and therefore new protein products within certain cell types. These novel protein products or peptides, which are expressed within the cells and presented by the major histocompatibility complex on the cell surface, have to be introduced to the immune system to avoid autoimmune reaction. Otherwise any tissue having a newly emerged peptide, which is represented on their cell surface, will be considered as foreign antigen and tissue/cells that are having such peptide will be attacked and destroyed by the immune system. Therefore, we propose that *de novo* evolved protein-coding genes should also be expressed within the thymus tissues to generate self-tolerance and avoid autoimmunity.
Based on the analysis from Thymus RNA seq data along with 10 other tissues within the phylogeny of Mus genus, we provide evidence that Thymus plays a very critical role in the evolution of metazoan by controlling birth of a new gene. Our results indicate a primary role for the Thymus controlling expression of all protein coding genes within both annotated Genic and Non Genic regions. The mechanisms may also be relevant for hybrid incompatibility effects between species and sub-species and thus also of relevance for speciation processes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML5

Evolutionary rescue from mutational meltdown

Claudia Bank [1,*]

[1]Instituto Gulbenkian de Ciencia, Oeiras, Portugal

Abstract: Mutagenic drugs are promising candidates for the treatment of various RNA virus infections. By increasing the mutation rate of the virus they lead to rapid accumulation of deleterious mutations, which is proposed to ultimately result in extinction as described by the theoretical concepts of mutational meltdown and lethal mutagenesis. However, the conditions and potential mechanisms of viral escape from the effects of mutagenic drugs have not been systematically explored. Using mutation-selection models, I investigate the population dynamics and genetics under high mutation rates and discuss the possibility of evolutionary rescue by means of a mutation rate modifier (i.e., evolution of resistance) versus mechanisms that rescue the population without modification of the mutation rate (i.e., evolution of tolerance). I show that extinction times are almost deterministic and that any rescue mechanism has to appear early to override the increasing mutation load. Importantly, there exist no robust experimentally assessable indicators to distinguish regimes in which evolutionary rescue is feasible from those in which extinction is unavoidable. This highlights the potential dangers of the use of mutagenic treatments, which are almost impossible to predict from experimental trials.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML14

Negative selection in humans and fruit flies involves synergistic epistasis

Mashaal Sohail [1 2 3,*], Olga Vakhrusheva [4 5], Jae Hoon Sul [6], Sara Pulit [7 8], Laurent Francioli [8], Leonard van den Berg [7], Jan Veldink [7], Paul de Bakker [8 9], Georgii Bazykin [4 5 10 11], Alexey Kondrashov [11 12], Shamil Sunyaev [2 3] and Genome of the Netherlands Consortium and Alzheimer's Disease Neuroimaging Initiative

[1]Systems Biology PhD Program, Harvard Medical School, [2]Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, [3]Department of Biomedical Informatics, Harvard Medical School, Boston, United States, [4]Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences, [5]Pirogov Russian National Research Medical University, Moscow, Russian Federation, [6]Department of Psychiatry and Biobehavioral Sciences, UCLA, Los Angeles, United States, [7]Department of Neurology and Neurosurgery, [8]Department of Genetics, [9]Department of Epidemiology, University Medical Center Utrecht, Utrecht, Netherlands, [10]Skolkovo Institute of Science and Technology, Skolkovo , [11]Department of Bioengineering and Bioinformatics, M.V. Lomonosov Moscow State University, Moscow, Russian Federation, [12]Department of Ecology and Evolutionary Biology, University of Michigan, Michigan, United States

Abstract: A long-standing puzzle in evolutionary genetics is the population survival under strong pressure of incessant deleterious mutations. Recent direct estimates of human and *D.melanogaster* genomic mutation rates suggest that these rates are high and likely incompatible with population survival under simple multiplicative models of negative selection. A proposed theoretical solution involves negative selection with synergistic, or narrowing, epistasis. We developed a statistical approach to test this conjecture directly in large sequencing datasets.

Negative selection with synergistic epistasis must produce negative linkage disequilibrium (LD) between deleterious alleles, and therefore, an under-dispersed distribution of the number of deleterious alleles in the genome. Indeed, we detected under-dispersion of the genomic number of rare Loss-of-Function (LoF) alleles in eight independent datasets from human and *D.melanogaster* populations of European and non-European ancestry. We also detected under-dispersion of the number of missense mutations in highly constrained genes in both organisms. Further, by conducting regression analyses and forward simulations, we ruled out other potential sources of negative LD such as random genetic drift and population structure.

Thus, empirical data suggests that ongoing selection against rare protein disrupting alleles is characterized by synergistic epistasis, which can explain how human and fly populations persist despite very high genomic deleterious mutation rates.

Expanded summary*: Little is known about the role of epistasis in fitness. Synergistic epistasis, whereby multiple deleterious genetic variants have a larger cost on fitness than expected from their multiplicative effects, plays an important role in evolutionary dynamics, e.g. by maintaining sexual reproduction and reducing the deleterious mutation load. Its prevalence in humans, however, still remains unknown.

We developed a statistical method to quantify the direction of epistasis in an organism's fitness using DNA sequencing data. Our method uses the distribution of the number of deleterious variants carried by an individual, or the deleterious mutation burden. We compute an interaction statistic *I*, the difference between the empirical variance of the deleterious mutation burden and the additive variance expected under no epistasis. Our test relies on an observed under-dispersion of the empirical variance due to "repulsion", or negative linkage disequilibrium (LD), between synergistically interacting deleterious variants.

We applied our method to six sequencing datasets – three European, Genome of the Netherlands, Alzheimer's Disease Neuroimaging Initiative, controls from a amyotrophic lateral sclerosis study, and three non-European populations from the 1000 genomes Phase I Project. We observe a significant under-dispersion signal in rare Loss-of-Function (nonsense and splice disrupting) variants across

human datasets. We also replicated our findings in the *Drosophila* Population Genomics Project and *D. melanogaster* Genetic Reference Panel. Further, we observe an under-dispersion signal in missense variants in highly constrained genes in both organisms. We ruled out genetic drift and population structure as sources of negative LD using regression models and forward simulations. To our knowledge, our results are the first to show that empirical data points towards the presence of synergistic, or narrowing, epistasis between deleterious variants involved in human fitness. Thirty years ago, Neel asked: "The amount of silent DNA is steadily shrinking. The question of how our species accommodates such [high deleterious] mutation rates is central to evolutionary thought." Indeed, an average human individual is expected to carry at least seven *de novo* deleterious mutations, which is inconsistent with the existence of the population at equilibrium under multiplicative selection. In a sense then, our finding that narrowing negative selection in humans is ongoing, is not unexpected.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-32

Estimating the ages of singletons and other rare variants

Alexander Platt*, Jody Hey [1]

[1]Center for Computational Genetics and Genomics, Temple University, Philadelphia, United States

Abstract: The ages of different variants segregating in the human population is a topic of considerable current interest. In a typical population genomic sample, however, a full half of the identified variants are present in only a single copy, and many of the variants found in any newly-sequenced individual will not have been observed in any previous sample. Existing methods that estimate ages based on allele frequencies can only assign all of these variants to a single, youngest, age class, and methods based on the decay of linkage disequilbrium or shared haplotypes surrounding rare alleles are inapplicable. This leaves this large class of variants almost completely uncharacterised, particularly those found outside of annotated functional regions of the genome.
There exists, however, real information in a population genomic sample that will allow us to estimate the ages of individual singleton alleles. In an infinite sites model where each allele has a unique origin, the mutation that created an allele found in a singleton must post-date the most recent common ancestor shared between the individual carrying the singleton allele and any other individual in the sample. We present both maximum likelihood and Bayesian estimators of the time since this common ancestor conditional on the observed maximum length of haplotype shared between the individual containing the singleton allele and any other individual in the sample. This estimator applies not just to singletons, but any rare allele. For alleles present in multiple copies we propose a composite likelihood estimate of the age of the most recent common ancestor shared between all of the individuals carrying the allele and any of the other individuals in the sample.
Armed with the joint spectrum of allele-age/allele-frequencies, a population geneticist will have a rich new tool with which to describe the action of natural selection on specific classes of rare variants.
There exists, however, real information in a population genomic sample that will allow us to estimate the ages of individual singleton alleles. In an infinite sites model where each allele has a unique origin, the mutation that created an allele found in a singleton must post-date the most recent common ancestor shared between the individual carrying the singleton allele and any other individual in the sample. We present both maximum likelihood and Bayesian estimators of the time since this common ancestor conditional on the observed maximum length of haplotype shared between the individual containing the singleton allele and any other individual in the sample. This estimator applies not just to singletons, but any rare allele. For alleles present in multiple copies we propose a composite likelihood estimate of the age of the most recent common ancestor shared between all of the individuals carrying the allele and any of the other individuals in the sample.
Armed with the joint spectrum of allele-age/allele-frequencies, a population geneticist will have a rich new tool with which to describe the action of natural selection on specific classes of rare variants.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-55

Lineage specific selection and mutation load accumulation in 32 species in the Brassicaceae

Tyler Kent [1,*], Stephen Wright [1], J. Chris Pires [2], Rod Wing [3], M. Eric Schranz [4], Thomas Mitchell-Olds [5], Detlef Weigel [6], Jeremy Schmutz [7], Adrian Platts [8], Patrick Edger [9], Jacob Washburn [2], Shengqiang Shu [10]

[1]EEB, University of Toronto, Toronto, Canada, [2]University of Missouri, Columbia, [3]University of Arizona, Tuscon, United States, [4]Wageningen University, WAGENINGEN, Netherlands, [5]Duke University, Durham, United States, [6]Max Planck Institute, Tübingen, Germany, [7]Joint Genome Institute, Walnut Creek, [8]NYU, New York, [9]Michigan State University, East Lansing, [10]Lawrence Berkeley National Lab, Berkeley, United States

Abstract: Although a substantial portion of most genomes is held under selective constraint, major evolutionary transitions, such as shifts to selfing and whole genome duplications, can allow new mutations at these sites to evade selection, and may provide the opportunity for deleterious mutation accumulation in that species.

We used whole genome sequences of 32 species in the Brassicaceae, encompassing three independent whole genome duplications, as well as multiple transitions to selfing, to investigate the accumulation of and variation in mutational load across lineages. This enabled us to create a new resolved phylogeny for the family and to obtain base pair resolution of constraint in each species using Genomic Evolutionary Rate Profiling (GERP) scores. We also used large population genetic datasets in both *C. grandiflora* and *A. thaliana* to obtain estimates of the distribution of fitness effects for bins of GERP scores in order to estimate the fitness costs of varying degrees of constraint loss. By correlating these fitness effects with GERP scores, we were able to get lineage-specific estimates of the fitness costs due to load for all 32 species, without the need for population genetic datasets for each species.

We present the most comprehensive view of constraint and mutational load in the Brassicaceae to date, with varying levels of fitness costs between lineages in relation to multiple whole genome duplications. We found the majority of mutations with large fitness costs to occur in coding regions of each genome, but with a substantial amount occurring in noncoding DNA.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML3

The influence of selective sweeps on the relationship between the efficiency of natural selection and genetic diversity.

Jenny James [1,*], Adam Eyre-Walker [1], David Castellano [2]

[1]University of Sussex, Brighton, United Kingdom, [2]Bioinformatics research centre, Aarhus, Denmark

Abstract: The efficiency of selection is expected to be greater in species that are more genetically diverse, because both of these factors depend on the effective population size. We have explored the relationship between a measure of the efficiency of selection, the ratio of nonsynonymous to synonymous nucleotide site diversity, and synonymous nucleotide site diversity, in two datasets: between mammalian species for mtDNA, and within the Drosophila nuclear genome. In both cases we observe a significant negative correlation. However, in both cases the slope of this relationship is significantly steeper than we would predict from estimates of the distribution of fitness effects inferred from the site frequency spectrum. This discrepancy may be due to the action of linked selection. We use SLiM to perform forward simulations to assess the impact of linked selection on neutral and deleterious diversity, and find that selective sweeps reduce genetic diversity for neutral mutations, but that the effect is attenuated for mutations subject to negative selection, to the extent that selective sweeps can leave the level of diversity at strongly selected sites unaffected. Thus, the actions of linked selection will result in an increase in the slope of the relationship between genetic diversity and the efficiency of selection. The results also suggest that selective sweeps cannot be simply characterised in terms of a reduction in effective population size for genetic variation that is common to all mutations, and they may have important implications for understanding the maintenance of genetic variation that is subject to selection.

Expanded summary*: N_e determines the level of neutral genetic diversity in a population, and will also determine the efficiency of natural selection acting in a population. This is because genetic drift is stronger in populations with a small N_e, and so a larger proportion of mutations will be effectively neutral (Charlesworth 2009; Kimura 1984). Therefore, because genetic diversity and the efficiency of selection are both influenced by N_e, they are expected to be correlated, a prediction which has been upheld in a number of studies (e.g. Galtier 2015; Piganeau & Eyre-Walker 2009). In this research, we were interested in exploring and quantifying this relationship.

This question is of biological importance for two main reasons. Firstly, the neutral genetic diversity of a population is often used as a proxy for N_e; however, the extent to which genetic diversity can be used as an estimator of selective constraint acting in a population is not well known, and may have important implication for a number of fields, including both molecular evolution and conservation genetics. Secondly, we can make specific predictions regarding the relationship between genetic diversity and the efficiency of selection: we expect the relationship between genetic diversity piS, and the efficiency of selection, as determined by piN/piS, to be loglinear, with a slope that is equal to the shape parameter of the gamma distribution of fitness effects (Welch et al. 2008). These predictions have not been previously tested.

Therefore, we developed a new method to explore the relationship between our measure of the efficiency of selection, the ratio of nonsynomous to synonymous nucleotide site diversity (piN/piS), and genetic diversity, measured as synonymous nucleotide site diversity (piS). Our method corrects for the statistical non-independence between these variables. We applied our method to two datasets: a between species dataset of mammalian mitochondrial DNA, and a within-species dataset using the *Drosophila melanogaster* nuclear genome, comparing between autosomal genes. In both datasets, we found a significant, negative loglinear relationship, however, surprisingly in both datasets the slope of the relationship was significantly steeper than that predicted by the shape parameter of the gamma distribution of fitness effects, as estimated from the site frequency spectrum.

We hypothesise that linked selection could be causing this discrepancy. In order to test this, we ran a number of forward simulations using the software package SLiM. In our study, we simulated a locus subject to adaptive evolution (in terms of varying frequencies of selective sweeps), linked to a locus containing sites under varying levels of negative selection. We found that while selective sweeps decreased the level of genetic variation for neutral and weakly deleterious variation, sweeps have far less effect on the diversity of more strongly deleterious variants, to the extent that levels of genetic diversity at those sites under the strongest negative selection are unaffected by selective sweeps. This can be thought of in terms of the equilibrium frequencies of different variants: deleterious genetic variants in a population are segregating at lower frequencies than neutral variants, and so will recover their equilibrium frequencies more rapidly after a sweep than will neutral variants, resulting in an increase in the relative frequency of deleterious variants (Marsden

et al. 2016; Messer & Petrov 2013). These simulations explain why the relationship between piN/piS and piS in both the mammalian mitochondrial dataset and Drosophila nuclear dataset is steeper than expected. In addition, they highlight how selective sweeps will not simply reduce Ne across a locus, irrespective of the type of sites at that locus. These results may help us to better understand levels of standing genetic variation within populations.

References

Charlesworth, B., 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature reviews. Genetics*, 10(3), pp.195–205.

Galtier, N., 2015. Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis. *PLoS genetics*, 12(1), p.e1005774.

Kimura, M., 1984. *The neutral theory of molecular evolution*, Cambridge University Press.

Marsden, C.D. et al., 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proceedings of the National Academy of Sciences*, 113(1), pp.152–157.

Messer, P.W. & Petrov, D.A., 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends in ecology & evolution*, 28, pp.659–69.

Piganeau, G. & Eyre-Walker, A., 2009. Evidence for variation in the effective population size of animal mitochondrial DNA. *PLoS ONE*, 4(2), p.e4396.

Welch, J.J., Eyre-Walker, A. & Waxman, D., 2008. Divergence and polymorphism under the nearly neutral theory of molecular evolution. *Journal of Molecular Evolution*, 67, pp.418–426.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-39

Meiotic Transgenerational Error Correction

Virgil Reese*

Abstract: Meiosis makes possible a mechanism by which sexually reproducing organisms can directly remove from their germ cells some inherited point mutations. When the close alignment and of homologous chromosomes in synapsis encounters any heterozygosity in a highly conserved (thus homozygous) DNA sequence, the presence of a likely point mutation is revealed. Because the organism has way no discern which of the two non-matching alleles to suspect, this information is of no immediate use. However, if the cell were to attach (at or near this site) an epigenetic "suspect-allele" tag, to both chromosomes, then when one of these chromosomes aligns with an new homolog after fertilization, this *suspect-allele* tag will give the new organism information it can use to initiate gene conversion that is biased against this initiating strand, and thus in favor of the unmarked (probable ancestral) allele from the new homologous chromosome. This repair process might conceivably operate in any of the new organism's cells, but a likely adaptive focus would be germline cells, perhaps occurring concurrently with the process of finding (and tagging for removal in a subsequent meiosis) additional isolated (not yet tagged) SNPs. Strong circumstantial evidence for the occurrence of this "Meiotic Trans-generational Error-Correction" (M-TEC) process will be offered.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-36

Evolvability phase diagrams yield long-term implications of fitness-mediated epistatic interactions

Matthew Melissa [1,*], Benjamin Good [2], Michael Desai [3]

[1]Physics, Massachussetts Institute of Technology, Cambridge, [2]Physics, University of California, Berkeley, Berkeley, [3]Organismic and Evolutionary Biology, Harvard University, Cambridge, United States

Abstract: The dynamics of asexual populations subject to both beneficial and deleterious mutations can be surprisingly complicated. In many biologically relevant parameter regimes, multiple mutations may segregate in a population at the same time, so that the eventual fate of a mutation depends not only on its fitness effect, but on the fitness of the genetic background it lands on. By appealing to a relation between coalescent timescales and fixation probabilities, an "evolvability phase diagram" is constructed, describing whether a population subject to mutations of a given rate and a given distribution of fitness effects will adapt, or decline, in fitness. We find that for a given size of beneficial fitness effects, and intermediate range of deleterious fitness effects correspond to fitness decline. This range depends on the population size as well as beneficial and deleterious mutation rates. Above a critical beneficial effect size, the population will adapt no matter what the size of deleterious effects. Given additional assumptions of the precise type of fitness-mediated epistatic interactions, these evolvability phase diagrams yield corresponding predictions regarding the existence and location of long-term evolutionary attractors in population-genetic parameter space.

Expanded summary*: Most work in theoretical population genetics assumes a simple set of rules that govern the evolutionary dynamics—for instance, a population might consist of a fixed number of individuals, subject to mutations which occur at a constant rate and have a particular effect on an individual's fitness, or growth rate. Fitness effects can be positive, negative, or neutral, and are typically assumed to be drawn from some known distribution. By making the assumption that each mutation can be treated independently, classical results have been obtained for how fast populations will evolve and what kind of genetic diversity can be expected.

It turns out that in some of the simplest real-world populations—say, a test-tube of yeast or E. coli.—these classical results have limited applicability, and the evolutionary dynamics are still not completely understood. In rapidly-mutating populations such as these, multiple mutations may segregate in a population at the same time, so that a mutation's eventual fate depends not only on its fitness effect, but also on the genetic background it lands on. A lineage might need to acquire several beneficial mutations in order to take over a population; alternatively, a lineage that initially increases in frequency might eventually be outcompeted by another lineage that acquires an exceptionally strong beneficial mutation. Treating these many-body effects has been challenging because the relevant birth-death process yields a set of stochastic, nonlinear and nonlocal partial differential equations.

For instance, consider a population subject to both beneficial and deleterious mutations, with each type of mutation occurring at a given rate and with a particular fitness effect. At what rate will the mean fitness of the population increase or decrease? Except in special cases, this question remains unanswered. I am working to answer this question, focusing in particular on whether the population will adapt, or decline in fitness. Can we gain an intuitive picture of the "evolvability phase diagram", describing which regions in parameter space correspond to adaptation, and which correspond to fitness decline? I have found that, somewhat unsurprisingly, increasing the effect size of beneficial mutations always results in faster adaptation. Somewhat less intuitively, we see that for a particular strength of beneficial mutations, an intermediate range of deleterious effect sizes brings about fitness decline—the deleterious mutations need to be strong enough to counteract the beneficial mutations, but weak enough not to be purged from the population immediately. If beneficial mutations are sufficiently strong, then the population will adapt no matter what the effect size of deleterious mutations. In a similar vein, we can also explore how evolvability depends on other factors such as the overall mutation rate and the shape of the distribution of fitness effects.

The model considered above neglects any interactions among mutations within an individual's genome: an individual's fitness is given by the sum of the fitness effects of their mutations. This is clearly biologically implausible, but is a natural first step at capturing the most basic evolutionary dynamics. During my PhD, I hope to make progress on extending these models to account for epistatic interactions—interactions in which the fitness effect of a new mutation depends on the mutations already present in the genome. As an individual acquires more mutations, the distribution of available fitness effects might change. A particularly tractable class of epistatic

interactions involves changes to the distribution of fitness effects that are mediated purely by an individual's fitness—for example, new beneficial mutations might yield diminishing returns as individuals get more fit. Using knowledge of the "evolvability phase diagram", I plan to investigate how the expected long-term flow of the population in parameter space depends on the precise model of fitness-mediated epistasis assumed. This approach is reasonable when changes to the distribution of fitness effects occur over long evolutionary timescales. A longer-term goal of my PhD is to investigate more general epistatic scenarios, in which the distribution of fitness effects might vary significantly over the course of fixation of a single mutant lineage.
A more fully developed conceptual framework describing evolutionary dynamics could be applied to diverse biological scenarios. For example, we could presumably forecast the onset of antibiotic resistance in bacterial populations, and how the time to develop resistance depends on the population size, drug concentration, or mutation rate. Other applications of this theoretical framework include understanding the impact of deleterious mutations on cancer progression and predicting dominant influenza strains that could be targeted by vaccines.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-38

Integrating ecology and evolution to study hypothetical dynamics of algal blooms and Muller's ratchet using Evolvix

Sarah Northey [1], Courtney Hove [1], Justine Kao [1], Jon Ide [1], Janel McKinney [1], Laurence Loewe [1,*]

[1]Laboratory of Genetics and Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, United States

Abstract: Algal blooms reoccur over various temporal and spatial scales. They occasionally carry toxins that can be very disruptive to their ecosystem. Therefore, algal blooms have been the subject of considerable research. Many algal blooms are governed by the availability of nutrients or other limiting growth factors. However, other limitations could exist. For example, algae are primary producers that can be subject to predation. This trophic relationship could result in typical predator-prey dynamics, where predation pressure keeps algal populations at levels below those supported by available nutrients. If a prey algae develops the ability to produce a toxin, that deters predators and thereby increases survival rates, then such algae might form blooms if supported by their environment. It is also clear that such algae will experience mutations that occasionally could knock-out DNA repair mechanisms, which could increase the probability of acquiring the necessary mutations for producing a predator deterring toxin.
Here we investigate the hypothetical scenario above. In our simulation model, we implement a sequence of steps that allows an asexual healthy algal population to escape predation pressure and form a bloom with the help of mutators. However, the bloom is then driven to extinction by the very cause of its initial success. The necessary steps are: (1) Loss of enough important DNA repair genes by random mutation. (2) The resulting increased mutation rate makes it more likely to acquire a mutation that alters algal metabolism to produce a toxin. (3) As a result, predators avoid toxic algae and provide them with a substantial growth advantage that can mask other deleterious mutational effects in the algal genome. (4) The continued lack of predation pressure caused by the toxin results in an algal bloom if sufficient nutrients are available and predation is strong enough on non-toxic algae. (5) The lack of recombination in these blooming mutators-algae inevitably causes a fast accumulation of slightly deleterious mutations as predicted by Muller's ratchet. (6) If such mutations accumulate fast enough they will eventually lead to the mutational meltdown of the toxic blooming algae, albeit without affecting the healthy algal population that remains under predation pressure. We simulate a corresponding model of algal blooms that integrates the ecological continuous time dynamics of a predator-prey system with the population genetics of a simplified Muller's ratchet model using the Evolvix modeling language. Even if our simulated model might be hypothetical, the need to simulate ecological and population dynamics and evolutionary genetics in the same integrate model is more wide spread.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML2

Patterns of deleterious variation within and between geographically diverse populations

Joshua Akey [1,*]

[1]Department of Genome Sciences, University of Washington, Seattle, United States

Abstract: The deluge of genome-scale sequence data now available in geographically diverse populations has provided considerable insights into patterns of genomic variation within and between populations. However, a coherent narrative about the characteristics, patterns, and consequences of deleterious mutations among individuals that have experienced different demographic histories has yet to emerge and many questions about deleterious variation remain. Here, I will present new analyses of deleterious protein-coding and regulatory mutations in tens of thousands of geographically diverse individuals. We show that demographic history and natural selection both influence patterns of deleterious variation, often in complicated ways, but whether such differences in patterns of deleterious influences mutational load, depends on the particular fitness model assumed. Moreover, using exome data from over 60,000 individuals we show a marked decrease in the average strength of selection acting on deleterious protein-coding variation over the past millennia. Finally, we show how archaic admixture influences the burden of deleterious mutations carried by individuals.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-34

African ROH Drive Enrichment of Deleterious Alleles in a Sample of Admixed Individuals

Zachary Szpiech*, Angel Mak, Marquitta White, Donglei Hu, Celeste Eng, Esteban Burchard, Ryan Hernandez

Abstract: Runs of homozygosity (ROH) are important genomic features that manifest when identical-by-descent haplotypes are inherited from parents. Their length distributions are informative about population history, and their genomic locations are useful for mapping recessive loci contributing to both Mendelian and complex disease risk. We have previously shown that ROH, and especially long ROH that are likely the result of recent parental relatedness, are enriched for homozygous deleterious coding variation in a worldwide sample of outbred individuals (Szpiech, et al. 2013). However, the distribution of ROH in admixed populations and their relationship to deleterious homozygous genotypes is understudied.

Here we analyze whole genome sequencing data from 1,484 individuals from African American, Puerto Rican, and Mexican American populations. These populations are three-way admixed between European, African, and Native American ancestries and provide an opportunity to study the distribution of deleterious alleles partitioned by local ancestry and ROH. We re-capitulate previous findings that long ROH are enriched for deleterious variation genome-wide. Then, partitioning by local ancestry, we compare the proportion of deleterious homozygotes in ROH comprised of single ancestry haplotypes to the proportion of benign homozygotes in those ROH. We find that ROH falling in African ancestry tracts are enriched the most followed by European and Native American ROH.

These results suggest that, while ROH on any haplotype background are associated with an inflation of deleterious homozygous variation, African haplotype backgrounds may play a particularly important role in the genetic architecture of complex diseases for admixed individuals, highlighting the need for further population genetic study of these populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-33

Differential Fitness Effects of Transitions and Transversions

Daniel Lyons [1,*], Adam Lauring [2]

[1]Ecology & Evolutionary Biology, [2]Microbiology, University of Michigan, Ann Arbor, United States

Abstract: Across most phyla, the substitution rates of transitions are higher than expected relative to those of transversions. One explanation is that natural selection favors transitions. This is based on the observation that non-synonymous transitions are more likely to conserve the biochemical properties of the original amino acid. Only recently has it become feasible to directly test this selective hypothesis by comparing the fitness values of a large number of transition and transversion mutations.

A recent analysis of available data in a beta-lactamase gene and in six viral taxa cast doubt on the selective hypothesis. This report failed to identify a biologically significant transition-transversion (TS-TV) fitness difference and argued that the supposed conservativeness of transitions has been based on arbitrary biochemical categories not necessarily reflective of selective constrains.

We performed a similar analysis on a genome-wide study of influenza. In our study, we find transversions to be significantly more detrimental than transitions. Then, using what we believe to be an improved statistical framework, we identify a similar fitness difference in HIV datasets used in the previous report and suggest that the other studies were inadequately powered. We confirm this fitness difference in four deep mutational scanning (DMS) datasets of influenza and HIV. We also show that three of the most commonly cited biochemical categories are predictive of fitness, supporting their use for this and other studies, such as the detection of positive selection and the origins of the genetic code. We conclude that selection likely contributes to the TS-TV substitution bias and that transitions are less detrimental in part due to their greater likelihood of conserving amino acid properties.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-37

Contribution of spontaneous mutations to the standing genetic variation in natural Arabidopsis thaliana populations

Mao-Lun Weng [1,*], Claude Becker [2], Matt Rutter [3], Detlef Weigel [4], Charles Fenster [1]

[1]South Dakota State University, Brookings, United States, [2]Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria, [3]College of Charleston, Charleston, United States, [4]Max Planck Institute for Developmental Biology, Tübingen, Germany

Abstract: Mutations are the ultimate source of all genetic variations, thus fundamental to the genesis of biodiversity. However direct estimates of the contribution of mutations to natural population, i.e., genetic variation underlying selection response, are limited. To estimate the mutational input to natural populations, we first analyzed the mutation rate and spectrum of mutations from *Arabidopsis thaliana* mutation accumulation (MA) lines at the 25^{th} generation, and then compared and contrasted the mutation profile against genetic variation observed in the 1001 Genomes Project, which characterized the global genetic variation pattern from 1135 *A. thaliana* accessions. The estimated mutation rate per site per generation was 5.65×10^{-9} and the INDELs mutation rate per site per generation was 1.31×10^{-9}. Among 1,694 SNPs identified in MA lines and 10,707,430 biallelic SNPs found in the 1001 Genome Project, 389 SNPs were at the same position suggesting that mutations are nonrandom and biallelic sites have a much higher mutation rate. Furthermore, among the mutations that occurred at the biallelic SNPs, 74% were the same nucleotide change suggesting that standing genetic variation observed in the natural populations is substantially shaped by mutational input. The contrast between MA lines and the 1001 Genomes Project revealed that the nucleotide changes were less skewed toward G:C to A:T in natural populations than in MA lines suggesting a GC-biased conversion that compensated the AT-biased mutation in the nature. The current study provides the most comprehensive estimate of mutation spectrum and mutational inputs for natural plant populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML4

The effect of background selection on mutational dynamics and diversity

Ivana Cvijovic [1,*], Benjamin Good, Michael Desai

[1]FAS Center for Systems Biology, Harvard University, Cambridge, United States

Abstract: Purifying selection shapes the genomic diversity of natural populations by reducing the diversity not only at directly selected sites, but also at linked neutral sites. This process is known as background selection, but its effect on the dynamics of mutations, or how it shapes diversity remain poorly understood. Here we analyze the forward-time dynamics of both neutral and deleterious mutations in populations experiencing background selection. We obtain analytical predictions for the distributions of mutational trajectories from which we calculate analytical predictions for diversity statistics. Our results offer intuition for the dynamics of polymorphisms under background selection, as well as offering an intuitive interpretation of the ways in which purifying selection shapes diversity. We find that purifying selection can lead to remarkably rich dynamics in populations with a wide distribution of fitnesses: neutral mutations at high frequencies exhibit sweep-like behaviors, and deleterious mutations can reach substantial frequencies, even though they are guaranteed to eventually go extinct. This peak frequency that deleterious mutations can reach depends strongly on both its own effect on fitness and the fitness of the ancestral background on which it arose. Beyond results on genomic diversity, our forward-time framework also allows the calculation of other quantities such as posterior distributions of the fitnesses of ancestral backgrounds of polymorphisms or their ages.

Expanded summary*:

Understanding the evolutionary pressures that shape natural populations is one of the major goals of evolutionary genetics. This is challenging, since evolution often happens on long timescales that we are not in a position to observe directly. Instead, we only have access to snapshots of the results of this process, encoded in the genomes of populations. In these snapshots, the genomic diversity of this population encodes a wealth of information about its evolutionary past. However, despite the growing abundance of data, interpreting this information has been challenging, since we simply lack predictions for the effects of selection on diversity in all but the simplest evolutionary scenarios. The main conceptual difficulty arises due to linkage: selection operates on an organismal level, and the fate of individual mutations depends not only on their merits alone, but also on the fitnesses of any linked polymorphisms. This is particularly obvious in asexually reproducing organisms, but continues to be relevant on short genomic distance scales in sexually reproducing organisms.

In the absence of theoretical models which would provide analytical predictions for the effects of linked selection on genomic diversity, we rely on whole-population whole-genome simulations to predict the effects of linked selection on diversity. These extremely computationally expensive methods have been complemented by more computationally efficient genealogical methods such as the coalescent, but these too remain largely numerical and offer limited intuition on how linked selection distorts genomic diversity. In particular, these simulations have given relatively well known yet surprising results. For instance, negative selection can in certain cases lead to signatures in genomic diversity that look very much like demographic expansions or even positive selection. Understanding these effects both intuitively and also quantitatively is essential to correct data interpretation.

We have developed a new theoretical approach for analyzing the effects of linked negative selection. Instead of starting backwards in time and attempting to model the genealogy, we instead model mutational dynamics directly and obtain an analytical description for the trajectories that both neutral and deleterious mutations take in the presence of negative selection. From these distributions of trajectories, we have been able to obtain analytical predictions for diversity statistics and how they depend on the parameters. This yields an intuitive explanation for the correspondence between the mutational dynamics and snapshots of diversity, and also offers an explanation for the contexts in which negative selection may look very much like a different evolutionary scenario.

Finally, our forward-time framework is also powerful in that it offers a way to calculate other quantities of interest beyond genetic diversity. In particular, it is possible to calculate posterior distributions of ancestral fitnesses for polymorphisms of specific frequencies in a way which accounts for mutations and any additional load accumulated on the background of that polymorphism.

Understanding the evolutionary pressures that shape natural populations is one of the major goals of evolutionary genetics. This is challenging, since evolution often happens on long timescales that we are not in a position to observe directly. Instead, we only have

access to snapshots of the results of this process, encoded in the genomes of populations. In these snapshots, the genomic diversity of this population encodes a wealth of information about its evolutionary past. However, despite the growing abundance of data, interpreting this information has been challenging, since we simply lack predictions for the effects of selection on diversity in all but the simplest evolutionary scenarios. The main conceptual difficulty arises due to linkage: selection operates on an organismal level, and the fate of individual mutations depends not only on their merits alone, but also on the fitnesses of any linked polymorphisms. This is particularly obvious in asexually reproducing organisms, but continues to be relevant on short genomic distance scales in sexually reproducing organisms.

In the absence of theoretical models which would provide analytical predictions for the effects of linked selection on genomic diversity, we rely on whole-population whole-genome simulations to predict the effects of linked selection on diversity. These extremely computationally expensive methods have been complemented by more computationally efficient genealogical methods such as the coalescent, but these too remain largely numerical and offer limited intuition on how linked selection distorts genomic diversity. In particular, these simulations have yielded relatively well known yet surprising results. For instance, negative selection can in certain cases lead to signatures in genomic diversity that look very much like demographic expansions or even positive selection. Understanding these effects both intuitively and also quantitatively is essential to correct data interpretation.

We have developed a new theoretical approach for analyzing the effects of linked negative selection. Instead of starting backwards in time and attempting to model the genealogy, we instead model mutational dynamics directly and obtain an analytical description for the trajectories that both neutral and deleterious mutations take in the presence of negative selection. From these distributions of trajectories, we have been able to obtain analytical predictions for diversity statistics and how they depend on the parameters. This yields an intuitive explanation for the correspondence between the mutational dynamics and snapshots of diversity, and also offers an explanation for the contexts in which negative selection may look very much like a different evolutionary scenario. Finally, our forward-time framework is also powerful in that it offers a way to calculate other quantities of interest beyond genetic diversity. In particular, it is possible to calculate posterior distributions of ancestral fitnesses for polymorphisms of specific frequencies in a way which accounts for mutations and any additional load accumulated on the background of that polymorphism.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-35

Species range limits, recombination rate variation, and chromosome inversions shape deleterious load in wild barley

Li Lei [1,*], Chaochih Liu [1], Paul J. Hoffman [1], Skylar R. Wyant [1], Ana M. Poets [1], Justin C. Fay [2], Peter L. Morrell [1] and Morrell lab

[1]Agronomy and Plant Genetics, UNIVERSITY OF MINNESOTA, St. Paul, [2]Genetics, Washington University, St. Louis, United States

Abstract: The distribution of deleterious variants across a species range and the genome of individuals is likely affected by effective population size (N_e) at multiple scales. For example, individuals in peripheral populations may be more prone to the accumulation of deleterious variants than individuals from centrally located populations because of frequent population founding and loss, and limited, unidirectional migration at the periphery of a species' range. Genomic regions with a low effective recombination rate also experience reduced N_e, resulting in a higher prevalence of deleterious variants. This effect could be amplified within structural variants such as inversions, which can result in dramatically suppressed recombination and high linkage disequilibrium (LD). We examine these effects in wild barley (*Hordeum vulgare* L. ssp. *spontaneum*), the progenitor of cultivated barley (*H. vulgare* L. ssp. *vulgare*). Using a combination of SNP genotyping and exome capture resequencing we make use of the draft barley reference genome to identify both the geographic and genomic distribution of deleterious mutations in 189 wild barley accessions. Our aim is to investigate how species range limits, recombination rate variation, and chromosome inversions shape the genomic and geographic distribution of deleterious variants in wild barley.

Expanded summary*: Naturally occurring diversity includes deleterious variants. Those variants are important to human health and also relevant to phenotypic variation in other species. Complementation of recessive deleterious variants between haplotypes is thought to be one of the primary mechanisms underlying heterosis (Charlesworth and Willis 2009). This suggests that identification of deleterious alleles may be applied to hybrid breeding strategies. Elevated proportions of deleterious variants relative to neutral ones in domesticated species suggest a cost of domestication (Cruz et al. 2008; Liu et al. 2017; Lu et al. 2006; Rodgers-Melnick et al. 2015). Understanding the origin and domestication of crop species can be enhanced by studying the genomic distribution and genetic contribution of deleterious variants. Deleterious variants can persist in populations owing to the limits of purifying selection. The potential for purifying selection to remove deleterious variants from populations is highly dependent on effective population size (N_e). When the product of N_e and selective coefficient (s) is small, less than 1, purifying selection is less powerful than genetic drift and deleterious variants tend to have a higher chance to be fixed in the population (Kimura et al., 1963). This is true at the species, population, and genome level. For example, N_e impacts how deleterious variants are distributed across the genome of individuals and also how deleterious variants are distributed across a species range.

The cause of species range limits are explained by several models, for example, recent arrival, metapopulation dynamics at the range periphery, source-sink dynamics, maladaptive gene flow and genetic constraints (Moeller et al., 2011). Regardless of which models, compared with core populations, peripheral populations exhibit increased genetic isolation and reductions in N_e due to limited and unidirectional migration (Hoffmann and Blows, 1994; Sexton et al., 2009). Henry et al. (2015) found that the accumulation of deleterious load will severely reduce the extent of a range across an environmental gradient. For species that are highly self-fertilizing, N_e in local populations could be extremely small, as populations could be founded by single individuals (Baker, 1955; Nordborg et al., 2002). This could lead to an excess of rare, potentially deleterious variants in local populations (Cummings and Clegg, 1998), particularly at the edges of the species range, and might also result in more pronounced effects of small N_e at the edges of a species range where gene flow from neighboring populations is reduced. However, there is currently limited empirical examination of the impact of species range limits on the accumulation of deleterious variants.

Genomic regions with a low effective recombination rate also experience reduced N_e, resulting in a higher prevalence of deleterious variants (Kono et al., 2016; Muller, 1964; Zhang et al., 2016). This effect could be amplified within structural variants such as inversions. Chromosomal inversions can be maintained by selection because they can prevent locally adaptive alleles at two or more loci from being lost due to migration (Kirkpatrick and Barton, 2006), and appear to play an important evolutionary role in a wide range of organisms (Ayala et al., 2011; Fang et al., 2012, 2014; Hoffmann and Rieseberg, 2008; Huynh et al., 2011; Lowry and Willis,

2010; Stevison et al., 2011). Inversions can lead to dramatically suppressed recombination and high linkage disequilibrium (LD). The suppression of recombination results in high levels of allele frequency differentiation between chromosomal arrangements, which can be observed as elevated F_{ST}. F_{ST} has been used to detect the presence of inverted regions in *Drosophila*, maize, and barley (Andolfatto et al., 2001; Fang et al., 2012, 2014). The suppression of recombination also leads to reduced N_e for the inverted regions of the genome. High LD could cause deleterious mutations to hitchhike along with advantageous mutations (Kirkpatrick, 2010). Therefore, inverted regions could bear a heavier mutation load.

Wild barley (*Hordeum vulgare* L. ssp. *spontaneum*), is the progenitor of cultivated barley (*H. vulgare* L. ssp. *vulgare*), a cultigen grown from the Equator to the Arctic Circle. The extensive range of the domesticate is particularly noteworthy because the wild progenitor occurs over a relatively narrow latitudinal range (Harlan and Zohary, 1966). Wild barley provides a single system in which we can investigate how species range limits, recombination rate variation, and chromosome inversions shape the genomic and geographic distribution of deleterious variants in wild barley. Fang, et al. 2014 reported two putative chromosomal structural variations in wild barley populations, and found that the genomic regions involved contributed disproportionally to population subdivision and environmental associations related to rainfall and temperature regime, thus potentially contributing to locally adaptive differences in populations. However, Fang et al. 2014 were not able to identify the size of the genomic regions where putative inversions occurred, owing to limits of the barley reference genome available at that time. Beyond that, it is unclear if these two inverted regions accumulate different mutation load compared with the rest of the genome.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML10

Mutation load across mating systems: how does load change and how do we best estimate it?

Kimberly Gilbert [1,*], Aneil Agrawal [1], Stephen Wright [1]

[1]Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada

Abstract: A large fraction of mutations occurring that impact fitness are known to be deleterious, contributing to what is termed the mutation load. Theory predicts that transitions from outcrossing to self-fertilizing mating systems or from sexual to asexual reproduction can have important effects on the efficacy of both positive and negative selection, with key consequences for mean fitness and mutation load. However empirically estimating mutation load and detecting fitness declines in populations is fraught with difficulty, making it challenging to test for the effects of mating system on mutation load. Several approaches are currently used, including counting deleterious mutations, applying GERP or similar constraint-based models to estimate how deleterious mutations are, or applying a DFE (distribution of fitness effects) estimation method to understand the severity of fitness effects across mutations. We simulate across this range of reproductive modes and compare estimates of mutation load to their true parameters in each simulation. We also simulate realistic demographic scenarios of population bottlenecks and evolutionary transitions across mating systems to understand changes in mutation load through time and the power our estimation approaches may have to detect load. We find that although changes in mean fitness and genetic load occur rapidly following these transitions, accurate empirical estimates of load are difficult and counting the total number of deleterious mutations in a population gives an incorrect inference of mutation load. Changes in homozygosity during transitions impact the efficacy of selection on mutations of varying effect sizes and dominance, highlighting the importance of accurate load measures.

Expanded summary*: Mutation is the ultimate source of new genetic variation, yet despite the benefits introduced by adaptive mutations, we know that out of the portion of mutations that impact fitness, most are deleterious. The strength at which selection acts to mitigate deleterious mutations therefore determines the ability of populations to reduce their mutation load and persist into the future. Theory shows that mating system and reproductive mode can alter the efficacy of both positive and negative selection in the genome. There is tremendous variation in reproductive mode in the natural world, with varying rates of outcrossing versus self-fertilization or asexual versus sexual reproduction. The key difference among these reproductive modes is the amount of recombination occurring each generation. With low recombination, we expect more linkage disequilibrium, reduced genetic diversity, and therefore lower effective population sizes, leading to the reduced efficacy of selection and mutations of larger effect to fall into the nearly neutral zone where selection is unable to purge them from the population. In addition to recombination rate differences, self-fertilization leads to increased inbreeding within populations, increasing homozygosity and exposing recessive, deleterious alleles. To understand how these processes impact mutation load and our current abilities to infer this load in natural populations, we simulate populations through time that undergo evolutionary transitions from an outcrossing, sexual population to either an asexual population or a self-fertilizing population, populations experiencing a bottleneck, or the combination of both of these events. This most closely resembles many plant species, and may for example be representative of invasive plant species that experience bottlenecks and increase their proportion of self-fertilization to spread into new habitat successfully.

Using the SLiM simulation program, we model realistic genomes within our populations, and apply mutation parameters to best match the current understandong in the field for the distribution of fitness effects of new mutations. From these simulations, we obtain our true, observed values of mutation load, calculated from known fitness values. To assess how empirical load estimation approaches perform, we then compare this true load to several estimations including raw counts of total mutations and substitution rates for deleterious mutations. We also apply the DFE-alpha method to our data to estimate the distribution of fitness effects, compare the proportion of mutations estimated per class of strength of selection to the true numbers of mutations known to be present in each of these classes throughout the simulations, and also estimate our load parameters within these selection strength classes to understand if constraint-based inferences such as GERP improve load estimations beyond simple counts. We find that empirical inferences do differ from the true load when comparing to the total count of mutations. Evolutionary transitions lead to drastic changes in mutation load over short timescales, more than population bottlenecks alone. Raw mutation counts particularly suffer across these scenarios due to the differences in efficacy of selection across reproductive modes in our populations, by not being able to account for the differences

in effect sizes of mutations present in populations. Our results highlight important points for understanding mutation load and improving our empirical methods for accurately inferring these changes in load through time and space. As genomic data continues to grow, making the proper inferences across species is vital to ensure the best understanding of mutational processes and their impact on population fitness and adaptation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-45

A comprehensive portrait of human somatic mosaicism

Selina Vattathil [1,*], Joshua Akey [1]

[1]Genome Sciences, University of Washington, Seattle, United States

Abstract: An adult human body contains on the order of 10^{15} cells, all of which can be traced back to a single zygote. The large number of cell divisions that occur during an individual's lifetime provides many opportunities for somatic mutations to arise and accumulate. Somatic mutations have been found to drive many forms of cancer and over thirty other human diseases, and likely play an important role in ageing phenotypes. Theoretical and empirical studies suggest that the burden of somatic mutation in humans is high, but few systematic and comprehensive studies have been conducted to date to characterize human somatic variability in apparently healthy individuals.

As part of the enhanced Genes, Tissues, and Expression (eGTEx) Project, we generated high-coverage (>150x) exome sequence data from over 220 samples obtained from 10 individuals (ranging from 19 to 32 tissues per individual), and identified somatic single-nucleotide substitutions and larger chromosomal imbalances (deletions, duplications, and copy-neutral loss of heterozygosity). The multi-tissue sampling design allows us to infer the developmental timing of mutations and test the relationship between somatic mutations and biologically importat variables such as sex, age, and tissue type. We anticipate our comprehensive catalog of somatic mutations will enable new insights into development and mutational processes and mechanisms, and provide guidance for interpreting mutations observed in cancer samples.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-44

Genetic load in cancer cells

Yuezheng Zhang, Yawei Li, Chung-I Wu, Hurng-Yi Wang, Xuemei Lu*

Abstract: Populations accumulate deleterious mutations that have yet to be removed by natural selection. This mutated portion of the population that does not contribute to its fitness is called the genetic load and is proportional to the deleterious mutation rate. Given the genomic instability of cancer cells, the genetic load could be substantial if the instability indeed leads to loss of cell fitness. We hypothesize that because aneuploidy and copy number variations (CNVs) are the most common forms of genome instability, their fitness consequences may determine the genetic load in cancer cells. However, since deleterious and many slightly deleterious mutations are hardly detected in empirical studies in cancers, the proportion and the extent of deleterious mutations have been ignored, which might be a key parameter to describe tumorigenesis. To test this hypothesis, we randomly selected single cells from a HeLa cell line and measured the cell growth rate, the genetic changes and the resulting variation in growth rate. We observed rapid generation of heterogeneity in the growth rate within the population, and that CNV accumulation decreased cell fitness systematically. The rate of CNV mutation was estimated by measuring the growth rate of two daughter cells from single proliferation and found that there is approximately 1 deleterious mutation in every 4 cell divisions. Therefore, due to a high, deleterious CNV mutation rate, tumor cells inevitably accumulate deleterious CNVs and a large percentage of tumor cells are genetically defective. Accordingly, we observed that the average growth rate of tumor cell populations decreased in the short term since defective cells accumulated in the population and the variation of cellular growth rates within the population increased. By modeling the process of mutation accumulation and measuring cell growth rate, we estimated that the deleterious mutation rate in HeLa cells is about 0.29 per cell division, and that HeLa cells reduce roughly 5%>6% of fitness for every cell division. Through simulating long-term cancer growth, we also found that this high genetic load lead to a very low survival rate of cancer clones. The observations of a high proliferation rate and high genetic load in this representative tumor cell line indicates a “high risk, high reward” evolution strategy for tumor cells and suggests that increasing the level of genomic instability may cause the meltdown of tumor cell population.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-42

Preliminary Investigation of the Somatic Mutation Rate in Daphnia pulex.

Stephan Baehr [1,*], Haimanti Ray [1], Michael Lynch [1]

[1]Indiana University, Bloomington, United States

Abstract:

Somatic cells imperfectly replicate their DNA during development and in stem cell niches, and their DNA is habitually damaged over time. These DNA mutations accumulate, and are presumed to contribute to somatic phenotypic change over the course of an organism's lifespan, for example age-related phenotypic decline. Describing the link between phenotype and genotype over time requires direct measurement of the somatic mutation rate *in vivo*. The somatic mutation rate has been measured in several model organisms indirectly, but the resolution remains poor. We here describe the theory, optimizations, and preliminary results of using Circle-seq to directly measure the somatic mutation rate in *Daphnia pulex*.

Expanded summary*:

The genomes of all organisms change over time, and describing the variables that shape the direction and rate of change are of fundamental importance to understanding life as we know it. Change ultimately derives from errors produced by biological processes. If errors are allowed to accumulate in the absence of purifying selection, the genotypes and phenotypes of living things increase in variance and decrease in overall fitness over time. The basal rates at which errors occur are therefore core variables of life. In the context of evolutionary time, heritable errors that affect characteristics reside in DNA. Rates of DNA error, or rather DNA mutation, have been of highest interest given DNA's heritability relative to other error types, such as those found in RNA transcription or protein translation. Substantial progress has been made in quantifying the organism/germline DNA mutation rate across the tree of life in the last 10 years, via the marriage of classical mutation accumulation experiments with the then-emergent technology of high-throughput sequencing. These experiments have led to a description of known variables that influence the evolution of the germline mutation rate, compiled into the drift-barrier hypothesis.

Even as the germline mutation rate is itself subject to evolutionary modification, so too is the somatic mutation rate, and direct damage by the latter may impose selection on the former. Despite sharing identical genes between the germline and the soma, somatic mutation rates are estimated to be between two-fold and two orders of magnitude higher than that of the germline. This result is predicted in evolutionary theory, in that somatic tissues are "disposable", as they do not directly contribute to future generations. Disposable tissues need not receive the same maintenance devoted to germline tissues, requiring only a phenotype sufficient to maximize germline fitness in a single generation. As a result, DNA mutation is implicated as a driver of cancer prevalence and the aging process itself. Data so far collected appear to support this conjecture: aging and cancerous tissues exhibit dramatically increased mutational loads relative to the germline, and the emergence of oncogenes solidified the contribution of DNA mutation to cancer incidence. Indirect measurements of somatic mutation rates have relied on *in vitro* methods, reporter constructs, or detection methods that have trouble distuinguishing sequencing errors from true mutations of low genomic frequency. Better characterizing the divergence between somatic mutation rate and the germline rate will define the extent to which evolution has optimized an intra-organismal mutation rate, and provide definition to the biological reality of aging somatic cells.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-49

Adaptation in small populations leads to the evolution of drift robustness and a reduced drift load

Thomas LaBar [1,*], Christoph Adami [1]

[1]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, United States

Abstract: Weakened selection leads to the fixation of slightly-deleterious mutations in small populations and a reduction in fitness known as the drift load. While it is known that populations with a high mutation rate can evolve mutational robustness to *de-novo* mutations, it is unknown if small populations can evolve robustness to the accumulation of slightly-deleterious mutations, or drift robustness. To test whether drift robustness and reduced drift loads can evolve in small populations, we performed digital experimental evolution with the Avida digital evolution software. We found that small populations evolve drift-robust genomes that limit the likelihood of slightly-deleterious mutations, and thus limit their potential drift load. Large populations evolve drift-fragile genomes that increase the likelihood of slightly-deleterious mutations and thus increase their potential drift load. This trend does not occur because drift-robust genotypes have a selective advantage over drift-fragile genotypes in small populations. Instead, drift robustness evolves because small populations cannot adapt by fixing small-effect beneficial mutations; they must adapt by fixing large-effect beneficial mutations. These results (confirmed with a mathematical model) indicate that population size determines the characteristics of a population's deleterious load: small populations will have large-effect deleterious mutations and large populations will have small-effect deleterious mutations.

Expanded summary*: Most *de-novo* mutations are deleterious and result in a reduction in mean fitness known as the mutational load. Research on mutational load has shown that populations with high mutation rates can evolve reduced mutational loads (mutational robustness). However, small populations face an additional burden: the drift load. In these small populations, weakened selection leads to the fixation of small-effect deleterious mutations. While the evolution of mutational robustness is well understood, it is unknown if small populations can evolve reduced drift loads, or drift robustness.

To test if small populations evolve drift robustness, we used the Avida digital experimental evolution system. We found that genotypes evolved in small populations display reduced drift loads (drift robustness) compared to genotypes evolved in large populations. Small populations evolve genetic architectures that limit slightly-deleterious mutations. Large populations instead evolve genetic architectures with a large likelihood of slightly-deleterious mutations. However, drift robustness did not evolve because drift-robust genotypes out-competed drift-fragile genotypes in small populations. Instead, drift robustness evolved because small populations cannot climb fitness peaks by fixing small-effect beneficial mutations (drift-fragile peaks). These populations must adapt by fixing large-effect beneficial mutations; it is precisely these peaks that display the drift-robust genetic architecture.

These results suggest a novel evolutionary trajectory that may shape genetic architecture and genetic variation. For instance, these results predict that strongly deleterious mutations cause the mutational load of human populations, as humans have historically had a low effective population size. The trend towards drift-robust genetic architecture can possibly explain patterns of deleterious genetic variation, such as the variation relevant to disease-associated mutations, in human populations. Our results may also explain genomic characteristics of pathogenic organisms, such as RNA viruses, that undergo strong bottlenecks upon infection. While the high mutation rates of RNA viruses imply that these organisms should evolve mutational robustness, the average mutation for RNA viruses is often strongly deleterious. Our results suggest that strongly-deleterious mutations in RNA viruses may occur because these viruses have evolved drift robustness in response to population bottlenecks.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML11

Genetic diversity and the efficacy of purifying selection across plant and animal species

Jun Chen [1], Sylvain Glémin [1], Martin Lascoux [1,*]

[1]Ecology and Genetics, Uppsala University, Uppsala, Sweden

Abstract: A central question in evolutionary biology is why some species have more genetic diversity than others and a no less important question is why selection efficacy varies among species. While these questions have started to be tackled in animals, they have not been addressed to the same extent in plants. Here we estimated nucleotide diversity at synonymous, π_S, and nonsynonymous sites, π_N, and a measure of the efficacy of selection, the ratio π_N/π_S, in 34 animal and 28 plant species using full genome data. We then evaluated the relationship of nucleotide diversity and selection efficacy with effective population size, the Distribution of Fitness effect and life history traits (LHT). In animals, our data confirm that longevity and propagule size are the variables that best explain the variation in π_S among species. In plants longevity also plays a major role as well as mating system. As predicted by the nearly neutral theory of molecular evolution the log of the ratio π_N/π_S decreased linearly with the log of π_S but the slope was weaker in plants than in animals. This appears to be due to a higher mutation rate in long lived plants, and the difference disappears when π_S is rescaled by the mutation rate. Differences in the distribution of fitness effect of new mutations also contributed to variation in π_N/π_S among species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML12

Draft masks the genomic signature of selection despite thousands of generations of adaptation in mutator bacteria

Alejandro Couce Iglesias [1], Larissa Viraphong [2], Thomas Hindre [2], Christoph Feinauer [3], Martin Weigt [3], Dominique Schneider [2], Olivier Tenaillon [1,*]

[1]IAME Infection, Antimicrobial Modelling Evolution, Inserm U1137, Inserm, Université Paris Diderot, Paris, [2]Laboratoire Technologies de l'Ingénierie Médicale et de la Complexité — Informatique, Mathématiques et Applications (TIMC-IMAG), Université Grenoble Alpes, Grenoble, [3]Computational and Quantitative Biology, Institut de Calcul et de la Simulation, CNRS, Université Pierre et Marie Curie, Paris, France

Abstract: Understanding the extreme degree of variation observed among bacterial genomes is still an unresolved challenge in evolutionary biology. Despite decades of debate among selectionist and neutralist viewpoints, no consensus has emerged as to which factors mainly drive genome evolution. A major confounding factor has been hypothesized to be the frequent variation in mutation rate over evolutionary history, and documented for several species. Mutation accumulation experiments with hypermutable bacteria have provided partial support to this hypothesis. These results have however been obtained under conditions of extremely weak natural selection, casting doubts about their validity to real-world circumstances. Here, we circumvent this limitation by analysing mutator strains from the Richard Lenski's Long-Term Evolution Experiment that have been experiencing adaptive evolution during tens of thousands of generations. We first developed a novel framework to identify the relative contribution of mutation versus selection for shaping genomic patterns. We then validated this framework with genomes evolved under different regimes, including high-mutation low-selection (mutation accumulation experiments) and low-mutation strong-selection conditions (E. coli natural isolates). Our results show that, despite tens of thousands of generations of adaptive evolution, mutator genomes show very few signals of selection. These results emphasize the importance of genomic draft, or how the strong linkage inside bacterial genomes limits the ability of selection to purge deleterious mutations. Overall, these results support the notion that past historical events play an underappreciated role in shaping current-day bacterial genomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML6

An upper limit on the functional fraction of the human genome

Dan Graur*

Abstract: For the human population to maintain a constant size from generation to generation, an increase in fecundity must compensate for the reduction in the mean fitness of the population caused by deleterious mutations. The required increase depends on the deleterious mutation rate and the number of sites in the genome that are functional. These dependencies and the fact that fecundity cannot be arbitrarily large, allow us to estimate an upper limit for the fraction of the human genome that can be functional. By estimating the fraction of deleterious mutation out of all mutations in known functional regions, we conclude that the fraction of the human genome that can be functional cannot exceed 25%.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-40

Abrogation of mitochondrial mutational load as a longevity mechanism in long lived Chiroptera

David Jebb [1,*], Nicole Foley [1], Sébastien Puechmaille [1 2], Roger Ransome [3], Gareth Jones [3], Gerald Kerth [2], Emma Teeling [1]

[1]University College Dublin, Dublin, Ireland, [2]Greifswald University, Greifswald, Germany, [3]University of Bristol, Bristol, United Kingdom

Abstract: The Free Radical Theory of Ageing (FRTA) proposes that the mitogenome accumulates mutations due to reactive oxygen species (ROS) produced during oxidative phosphorylation. This age-related increase in mitochondrial mutational load (MML) in turn leads to the progressive ageing phenotype. MML, measured as the number of heteroplasmic sites in the mitogenome, has been found to increase with age in humans in many tissues. However, little is known about the relationship between MML and age in other mammals Bats are the only mammals capable of true powered flight. Concomitant with flight is an increased metabolic rate to fuel this energy intensive form of locomotion. However, bats exhibit exceptional longecvity defying the FRTA and rendering bats a unique model for the study of mammalian ageing. In this study, we investigated MML in 3 long-lived bats species, *Myotis myotis* (max. lifespan 37.1yrs), *M. bechsteinii* (21yrs) and *Rhinolophus ferrumequinum* (30.5yrs) using targeted, deep sequencing of mitogenomes at the population level. Our results show all the bat species exhibit similar levels of MML to each other and humans. GC-TA transversions, characteristic of oxidative stress, were not the primary source of mutations in any species, though the bats exhibited higher levels than reported in humans. Downstream analysis of protein coding mutations in *M. myotis* suggests positive selection acting on specific mutations, possibly selecting for more efficient mitochondrial haplotypes. Our findings suggest bats defy the FRTA by abrogating mitochondrial mutational load, and that intra-individual selection may act to produce the efficient mitochondria necessary for flight and longevity.

Expanded summary*: It is predicted that a child born in the western world today may live to be as old as 150 years of age. The World Health Organisation predicts the number of people over 60 years old will double before 2050 and while the number of those over 80 will more than triple. As people are living longer, the incidence of age related diseases is increasing. Advantages to "long-life" are lost without "long-health". We believe studying the evolution of long-health in exceptional mammals may provide a novel insight into human ageing.

Chiroptera, the bats, are exceptionally long lived mammals, given their body size and metabolic rate. After correcting for body size, metabolic rate and other ecological factors only 19 species live longer than humans, 18 of which are bats (the other is the naked mole rat). Bats are the only mammals capable of true powered flight. Concomitant with flight is an increased metabolic rate to fuel this energy intensive form of locomotion. Exceptional longevity in bats is counter to the Rate of Living Hypothesis (RLH) and the connected Free Radical Theory of Ageing (FRTA).

The RLH spawned from the observation that larger mammals tended to live longer than smaller mammals. As size is also correlated with metabolic rate, the RLH proposes that a high metabolic rate leads to a short lifespan, with slower metabolic rates leading to longer life. This idea was given a mechanism with the FRTA, which posited that through normal mitochondrial activity, reactive oxygen species (ROS) are generated leading to oxidative damage of biomolecules. Damaged mitochondria in turn produce more ROS leading to a vicious cycle and the familiar, gradual, progressive physiological decline associated with age. Central to the FRTA are mitochondrial DNA mutations, and mitochondrial mutational load (MML). Heteroplasmy is the presence of multiple, non-identical mitogenome within an individual. MML, measured as the number of heteroplasmic sites within an individual, has been shown to increase significantly with age in multiple studies of humans, in multiple tissues. MML in children has also been shown to increase with mother's age at conception.

Under the tenets of the RLH and FRTA, Chiroptera with their small size should be short lived. Even stranger, given their increased metabolic rate due to flight their lifespan should be decreased even further. The exceptional longevity observed in Chiroptera defies the FRTA and RLH.

In this study, in order to understand the unique longevity observed in Chiroptera, we deep sequenced (~3500X) whole mitogenomes from populations of 3 long lived at species *Myotis myotis* (adult weight: 28.55g; max. lifespan: 37.1yrs), *M. bechsteinii* (10.5g, 21yrs)

and *Rhinolophus ferrumequinum* (22.88g, 30.5yrs). The average MML was 1-3 sites for the bat species. This is similar to previously published estimates in humans. The vast majority of mutations were present at less than 5% frequency. The primary class of mutation were transitions (>60%), however oxidative transversions, GC→TA, were present at a much higher level than in humans. MML increased in *M. myotis* with age, at a higher rate than in humans. However, given a maximum lifespan of ~120 years in humans, both rates amount to a similar increase in heteroplasmy over each species' lifespan. Analysis of protein coding mutations found that mutations present at high frequency were enriched for nonsynonymous mutations. This may suggest positive selection acting at specific sites.
In conclusion, for the first time we have measured MML in three long lived bat species. We posit that bats abrogate MML reducing the effect of ROS allowing their exceptional longevity in spite of their increased metabolic rate, possible through stringent mitochondrial quality control. Intra-individual selection may also act to produce the efficient mitochondria necessary for flight and longevity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-41

Evaluating how differences in demography and lifestyle affect selection efficacy and mutational load

Marie Lopez [1,*], Athanasios Kousathanas [1], Luis B. Barreiro [2], George H. Perry [3], Hélène Quach [1], Christine Harmant [1], Evelyne Heyer [4], Paul Verdu [4], Etienne Patin [1], Lluis Quintana-Murci [1]

[1]Human Evolutionary Genetics Unit, CNRS URA 3012, Institut Pasteur, Paris, France, [2]Université de Montréal, Montréal, Canada, Montréal, Canada, [3]Pennsylvania State University, University Park, PA 16802, USA, State College, United States, [4]CNRS, MNHN, Université Paris Diderot, Paris, France, Paris, France

Abstract: Understanding how recent changes in demographic regimes impact selection efficacy and thus the ability to purge deleterious variants remains a major, though controversial, question in evolutionary biology. To gain novel insight into the factors influencing the mutational load, we reconstructed the demographic history of sub-Saharan African populations that strongly differ in their historical modes of subsistence, and evaluated the efficiency of purifying selection in removing nonsynonymous variants. To do so, we generated 300 high-coverage whole-exome sequences, leading to the identification of 488,653 SNPs, in various populations of rainforest hunter-gatherers and neighboring farmers from the western and eastern parts of the Central African belt. We showed, via model-based inference of population size changes, that the effective size of agriculturalists has doubled in the last 10,000 years while rainforest hunter-gatherers have experienced a 75% population contraction. We next tested for differences in selection efficacy among populations by inferring their distributions of fitness effects of new nonsynonymous mutations (DFE), and found no significant differences in the DFE of rainforest hunter-gatherers and farmers. The empirical quantification of the individual mutational load further supported this finding, as no significant differences were detected in the mean number of deleterious alleles per population, under a hypothesis of semidominance. Despite the strong differences in recent demographic history and historical lifestyle of our population setting, our results collectively show that no major differences exist in selection efficacy between populations, providing empirical support to theoretical studies suggesting that recent demographic history had limited impact on selection in humans.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-43

Impact of recombination on the base composition of Bacteria and Archaea

Louis-Marie Bobay [1,*], Howard Ochman [1]

[1]Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: The mutational process in bacteria is biased toward A and T, and most species are GC-rich relative to the mutational input to their genome. It has been proposed that the shift in base composition is an adaptive process—that natural selection operates to increase GC-contents—and there is experimental evidence that bacterial strains with GC-rich genomes have higher fitness than those strains with AT-rich genomes. Alternatively, a non-adaptive process, GC-biased gene conversion (gBGC), could also increase the GC-content of DNA due to the mechanistic bias of gene conversion events during recombination. To determine what role recombination plays in the base composition of bacterial genomes, we compared the spectrum of nucleotide polymorphisms introduced by recombination in all microbial species represented by large numbers of sequenced strains. We found that recombinant alleles are consistently biased toward A and T, and that the magnitude of AT-bias introduced by recombination is similar to that of mutations. These results indicate that recombination alone, without the intervention of selection, is unlikely to counteract the AT-enrichment of bacterial genomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-50

Under pressure: deleterious alleles, haploid selection and climate affecting Scots pine molecular diversity

Tanja Pyhäjärvi [1 2,*], Jaakko Tyrmi [1], Sami Saarenpää [1], Tuomas Hämälä [1 2], Outi Savolainen [1 2]

[1]Ecology and Genetics, [2]Biocenter Oulu, University of Oulu, Oulu, Finland

Abstract: Scots pine (*Pinus sylvestris*) is a widely distributed conifer inhabiting a variety of environmental conditions. For example, growing season length, day length and amount of precipitation are very different across its European distribution. Therefore, genetic variation governing its adaptation is under strong natural selection.
In addition to selection due to environment, some Scots pine genetic variation is under strong selection for intrinsic reasons. Scots pine, like other conifers, has a short but important haploid life stage. In contrast to flowering plants, megagametophyte (endosperm) tissue within seed is haploid in conifers. Therefore, genes expressed in this haploid tissue should be under stronger selection pressure, as recessive alleles are revealed to selection at the haploid stage.
Finally, it has long been known that Scots pine has a high number of lethal equivalents segregating in natural populations. However, the exact distribution of the deleterious variation along genome and its effect on surrounding genetic variation is not known.
We have used exome capture and sequencing to characterize genetic variation in populations across two different transects in Europe. This data is used to identify polymorphism underlying environmental adaptation. The same dataset is also be used to identify deleterious alleles and their effect on genetic diversity. Further, RNAseq has been used to identify endosperm-specifically expressed genes. By combining the exome sequence and RNAseq data, we also investigate the role of haploid selection in Scots pine. The goal is to understand the relative roles of haploid selection, deleterious alleles and adaptation affecting polymorphism along the genome.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-51

Accumulation of mutations in experimental evolution of basidiomycete fungus Schizophyllum commune

Aleksandra Bezmenova [1,*], Alexey Penin [1 2], Elena Zvyagina [3], Artem Kasianov [1 2], Tatiana Neretina [1], Georgii Bazykin [1 2 4], Alexey Kondrashov [5]

[1]Lomonosov Moscow State University, [2]Institute for Information Transmission Problems of RAS, Moscow, [3]Yugansky Nature Reserve, Ugut, [4]Skolkovo Institute of Science and Technology, Skolkovo, Russian Federation, [5]University of Michigan, Ann Arbor, United States

Abstract: Basidiomycete fungus *Schizophyllum commune* is a unique organism with the highest genetic diversity among studied species, and a high mutation rate of $2.0*10^{-8}$ substitutions per nucleotide per generation [1]. The life cycle of *S. commune* includes a mononuclear haploid stage called monospore culture, which originates from a single spore and can be cultivated on solid medium. This makes *S. commune* a promising object to study the mutation process during vegetative growth of the fungus. We designed an experiment which allows us to cultivate monospore cultures of *S. commune* in long tubes with fixed diameter of the section for long period of time. We have been cultivating two lines started from one monospore culture in very thin tubes, insuring small population sizes, for about a year, and collected three samples of the mycelium from each line – after about 15, 30 and 45 cm of growth, which corresponds to ~2000, 4000 and 6000 cell divisions. Sequencing of these samples revealed a provisional high rate of accumulation of mutations during vegetative growth, although variation is maintained even at such low population sizes.

1. Baranova A et al. (2015) Extraordinary Genetic Diversity in a Wood Decay Mushroom. Mol. Biol. Evol. 32(10):2775–2783 doi:10.1093/molbev/msv153

Statement: I am a sixth year student of Lomonosov Moscow State University, Faculty of Bioingeneering and Bioinformatics. I am currently working in the laboratory of Evolutionary Genomics, headed by Prof. Alexey S. Kondrashov. I am carrying out an experiment on the long-term cultivation of basidiomycete fungi *Schizophyllum commune* designed to study the process of mutation accumulation. I would like to take part in the SMBE meeting as my first international conference. I hope to share my results with the international scientific community, practice my presentation skills and improve my English.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML1

The impact of demographic history on genetic load

Kirk Lohmueller [1,*], Bernard Kim [1], Clare Marsden [1], Jazlyn Mooney [1], Jesse Garcia [1], Jacqueline Robinson [1], Christian Huber [1]

[1]Department of Ecology and Evolutionary Biology, UCLA, Los Angeles, United States

Abstract: The fate of a deleterious mutation in a population depends on its effect on fitness and the demography of the population. Theory has suggested that in small populations, deleterious mutations can accumulate and become fixed, leading to an increase in genetic load. However, the extent to which non-equilibrium demographic forces impact the genetic load remains unclear. In this talk, I will present our recent empirical and simulation-based work exploring how demography impacts the genetic load. First, I will discuss how the bottlenecks associated with dog domestication have led to a 1-3% increase in the additive genetic load in dogs relative to wolves. Second, I will present the results of a simulation study examining the effects of admixture on genetic load. We find that substantial amounts of admixture (~25%) from a large population into a smaller population can substantially reduce the additive genetic load of the smaller population. However, without ongoing gene flow, the reduction in load is temporary with increasing load over time post-admixture. Conversely, smaller amounts of admixture (~1%) have little effect on the load of the small population. Lastly, I will explore the effect that recent inbreeding has on patterns of deleterious variants with different dominance coefficients and selective effects. In particular, simulations suggest that recent inbreeding reduces the number of variants within a run of homozygosity, but increases the number of homozygous derived deleterious variants within a run. Therefore, we found little change in the total number of derived alleles within a run compared to outside a run. Further, recent inbreeding can lead to an appreciable increase in genetic load of recessive mutations due to inbreeding depression. I will examine the aforementioned predications using genomic data from isolated human populations as well as in the context of non-human species undergoing population decline.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML8

Estimating the selective effect of heterozygous protein truncating variants from human exome data

Donate Weghorn [1,*], Christopher Cassa [1], Daniel Balick [1], Daniel Jordan [2], David Beier [3], Shamil Sunyaev [4]

[1]Brigham and Women's Hospital / Harvard Medical School, Boston, MA, [2]Icahn School of Medicine at Mount Sinai, New York, NY, [3]University of Washington School of Medicine, Seattle, WA, [4]Brigham and Women's Hospital / Harvard Medical School, Boston, MA, United States

Abstract: The dispensability of individual genes for viability has interested generations of geneticists. For some genes it is essential to maintain two functional chromosomal copies, while other genes may tolerate the loss of one or both copies. Exome sequence data from 60,706 individuals provide sufficient observations of rare protein truncating variants (PTVs) to make genome-wide estimates of selection against heterozygous loss of gene function. The cumulative frequency of rare deleterious PTVs is primarily determined by the balance between incoming mutations and purifying selection rather than genetic drift. This enables the estimation of the genome-wide distribution of selection coefficients for heterozygous PTVs, using a hierarchical model to fit the observed distribution of PTV counts. The inferred distribution of selective effects allows for a Bayesian estimation of the heterozygous selection coefficient for each individual gene. The estimated strength of selection can help discriminate the severity, age of onset, and mode of inheritance in Mendelian exome sequencing cases. We find that genes under the strongest selection are enriched in embryonic lethal mouse knockouts, putatively cell-essential genes inferred from human tumor cells, Mendelian disease genes, and key developmental pathways.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML9

Presence and ecological effect of a raccoon rabies expansion load

Olivia Yvellez [1,*], Eddie Zhao [1], Hannah Trewby [2], Roman Biek [2], Leslie Real [3], Katia Koelle [1]

[1]Biology , Duke University, Durham, United States, [2]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, United Kingdom, [3]Biology , Emory University , Atlanta , United States

Abstract: During a range expansion, sublethal deleterious mutations have the potential to 'gene surf' along a population's invasion front, resulting in an expansion load. Here, we analyze viral genetic data from a well-documented range expansion of raccoon rabies virus (RRV) in the eastern US to determine whether this population provides an empirical example of an expansion load. RRV provides an ideal system for documenting an expansion load because the virus is haploid, has a high mutation rate, and has been sampled over both space and time. Using model simulations, we first show that we expect a rabies expansion load to occur over a broad range of mutation fitness costs. We further show that this expansion load should slow the rate at which rabies spreads spatially and decrease overall rabies incidence. Based on these results, we performed a phylogenetic analysis of 47 partial G/partial N sequences sampled over 29 years, finding evidence for 7 spatially distinct lineages, consistent with a previous analysis. In the second northeast (NE2) lineage of RRV, we find preliminary evidence for an expansion load, based on elevated dN/dS ratios on the internal branches leading up to the viral samples isolated during the early epizootic expansion wave. Further, computational software predicts that the identified nonsynonymous mutations are likely to be deleterious based on their destabilizing effects. Beyond providing an empirical example of an expansion load, this work points towards the need for considering non-neutral evolutionary dynamics in understanding the spatio-temporal patterns of viral infectious disease spread.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

POB-47

Investigating transmission of mitochondrial DNA heteroplasmies across multi-generation pedigrees

Arslan Zaidi [1,*], Shu-Wei Su [1], Boris Rebolledo-Jaramillo [2], Jessica Beiler [3], Ian Paul [3], Peter Wilton [4], Anton Nekrutenko [2], Rasmus Nielsen [4], Kateryna Makova [1]

[1]Biology, [2]Biochemistry and Molecular Biology, [3]Department of Pediatrics, College of Medicine, Pennsylvania State University, University Park, [4]Department of Integrative Biology, University of California, Berkeley, Berkeley, United States

Abstract: Mutations in mitochondrial DNA (mtDNA) lead to heteroplasmies, the presence of more than one allele at a locus in an individual. Previous studies have shown that mtDNA undergoes a bottleneck during oogenesis. This reduces the mutation load in mtDNA, which has a higher mutation rate compared to the nuclear genome, and is non-recombining. However, a severe bottleneck could also lead to fixation, in children, of deleterious variants that are present at non-pathogenic frequencies in the mother's germline. Thus, it is clinically relevant to investigate mtDNA transmissions from mother to child, and to estimate the germline bottleneck size. We studied heteroplasmies in high-coverage mitochondrial genome sequences (>8,000x) from two tissues (buccal and blood) of 363 individuals from 101 multi-generational families, including 212 mother-child transmissions. We identified 769 heteroplasmies (with minor allele frequency above 1%), resulting in an average estimate of two heteroplasmies per person. There are fewer heteroplasmies in protein-coding regions than expected by chance, in agreement with purifying selection (p-value=5.02e-46). The number of heteroplasmies carried by a child is significantly associated with with mother's age at the time of giving birth (p-value=0.0136), suggesting that older mothers transmit more mutations to their children. Using the Wright-Fisher model, we estimated the median bottleneck size to be 34 segregating units (IQR:9.47-82.07). Additionally, we explore how rapid mtDNA replication after the germline bottleneck affects estimates of the bottleneck size. Our results demonstrate the effects of maternal age and germline bottleneck on mtDNA mutational load.

Expanded summary*: Mitochondria, because of their role in energy production and metabolism, are essential to cellular function. Impaired mitochondrial function, which may be due to mutations in mitochondrial DNA (mtDNA), is known to be associated with over 200 diseases. These mutations can arise de novo in the somatic tissues of an individual, or can be inherited from the germline of the mother, leading to more than one mtDNA variant in the same individual, a phenomenon called heteroplasmy. This concept is central to our understanding of mitochondrial disease etiology. Pathogenic mitochondrial mutations, if present over a threshold frequency within the cells of an individual, can lead to mitochondrial disease, the severity of which depends on the type of mutation and level of heteroplasmy. These mutations, if inherited from the mother, can rise to appreciable frequencies in children by genetic drift because of the germline bottleneck. This raises a number of important questions: How common are heteroplasmies? What proportion of these are due to pathogenic mutations? Do older mothers transmit more mutations? How severe is the germline bottleneck? Does its size vary among individuals? The goal of this study is to answer these questions by studying the transmission of heteroplasmies between mothers and their children.

To achieve this goal, we sequenced the mitochondrial genome at high-depth (>8,000x) of two tissues (blood and buccal) of 363 individuals from 101 families. Most of these families comprise one mother and at least two children, and some are multi-generational. In total, we recorded 212 mother-child transmissions. We detected 769 heteroplasmies across all individuals, based on a minor allele frequency threshold of 1%, leading to a mean estimate of 2 heteroplasmies per individual. These heteroplasmies are non-uniformly distributed along the length of the mtDNA molecule, with fewer than expected heteroplasmies in protein-coding regions (p-value=5.02e-46), suggesting purifying selection in these regions. We found a significant positive correlation

between the number of heteroplasmies carried by a child and the age of the mother at the time of giving birth (Poisson regression: t = 2.47, p-value = 0.0136), indicating that older mothers transmit more mutations to their children than younger mothers. In order to estimate the size of the germline bottleneck, we compared the heteroplasmy level between each mother and her children. The premise being that a severe bottleneck will lead to large fluctuations in heteroplasmy level between mother and child as a result of genetic drift. Based on a comparison of 66 mothers, who carry germline heteroplasmies, and 135 children, we estimated the germline bottleneck size to be 34 segregating units of mtDNA (IQR: 9.47 – 82.07), which is in agreement with previous estimates in humans. That this bottleneck is severe is evident in the observation that heteroplasmy levels are more correlated between tissues of the same individual (blood vs buccal: r_2 = 0.92 in mothers, r_2 in children) than between a mother and her child (mother vs child: r_2 = 0.59 for blood, r_2 = 0.53 for buccal). Currently we are investigating the effect of clonal expansion of mtDNA in the germline on estimates of the bottleneck size.
By providing an accurate estimate of the germline bottleneck size, this study will shed light on the risk of inheriting disease-causing mitochondrial mutations, which may persist in mothers at low frequencies. This information will be useful for genetic counseling and treatment of mitochondrial diseases.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML13

Prevalence of recessive selection confounds estimates of the mutation load in humans

Daniel J. Balick [1,*], Daniel Jordan [2], Shamil Sunyaev [3], Ron Do [2]

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School , Boston, MA, [2]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, [3]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Abstract: The mutation load is perhaps the simplest observable of natural selection on the population level. However, it remains amongst the hardest to measure in long-lived diploid organisms, owing to the unknown role and prevalence of genetic dominance. In humans, diploid selection coefficients cannot be measured experimentally, unlike many model organisms, such that inference techniques from both population genetic and phylogenetic data must be used to identify the distribution of selection and dominance coefficients. Little progress has been made to this end, as recessive variation is fundamentally more sensitive to non-equilibrium demography, leading to degeneracies in most widely used statistics. The majority of statistics used to identify genes under selection cannot distinguish recessive selection from additive variation under weaker selection due to the prevalence of neutral heterozygotes. For simplicity, many studies either explicitly assume additivity to limit the number of inference parameters, or implicitly assume additivity by using the mutation burden (the sum of frequencies) as a surrogate for the mutation load, avoiding the effects of diploidy completely. Since mutations in model organisms, including flies and yeast, are believed to exhibit an inverse correlation between selection strengths and dominance coefficients, this leads to a potentially substantial underestimation of the average strength of natural selection in humans, as a class of strongly selected variation remains undetected. The diploid mutation load, a function of the weighted average of the heterozygote and homozygote selection strengths, is thus confounded to an unknown extent, as both the relevant weightings and selection coefficients are altered by recessivity.

We show that four common population-based measures of natural selection, the number of segregating sites, the number of rare sites, the heterozygosity, and the mutation burden, are unable to accurately infer recessive selection. We detail specific biases of each, and that they are most sensitive to recessive variation in distinct regimes. Additionally, we show that a maximum likelihood inference from the site frequency spectrum (SFS) suffers from similar inaccuracies. We describe the stability of each measure in the presence of non-equilibrium demography, with particular attention to robustness to demographic inaccuracies. Inference from these statistics, per-gene or in aggregate, is generally used to estimate the mutation load.

To address the lack of statistics sensitive to dominance coefficients, we introduce a metric that distinguishes between recessive and additive strong selection, weak selection, and neutrality on the per-gene level. We combine two summary statistics, the mutation burden and a logarithmic transformation of the SFS, to simultaneously identify an average selection and dominance coefficient and eliminate the degeneracy seen in other statistics. We apply our method to infer per-gene scores from roughly 35K European exomes in ExAC. We validate our results using recessive disease genes, haploinsufficiency, and autozygosity data from consanguineous populations, finding highly significant enrichments in the expected directions. We find that 30-40% genes contain predicted protein damaging sites under strong purifying recessive selection ($h \leq 0.1$), which suggests the mutation load in humans is likely to be dramatically altered, and must be revised to include the effects of diploid selection.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational load

OW-ML7

The population genetics of recessive, lethal mutations

Carlos Eduardo G. Amorim*, Ziyue Gao, Zachary Baker, José Francisco Diesel, Yuval B. Simons, Imran S. Haque, Joseph Pickrell, Molly Przeworski

Abstract: What determines the frequencies of deleterious mutations? To begin to answer this question, we focus on mutations reported to cause recessive, lethal Mendelian diseases. We first review long-standing, analytic models of mutation-selection balance and compare them to simulations of purifying selection in a more realistic demographic setting. We further characterize the empirical distribution of strongly deleterious mutations in human populations, showing that they are not restricted to low frequencies, and test how well these models fit the allele frequencies estimated from 33,370 European individuals. We find that, for a highly mutable type of mutation, predictions fit observed frequencies well. For less mutable types, however, predictions tend to greatly under-estimate observed frequencies; this discrepancy is even larger when subtle fitness effects in heterozygotes or lethal compound heterozygotes are taken into account. We discuss possible explanations for this discrepancy (e.g. balancing selection, genetic modifiers) and point to a complication that, to our knowledge, is not widely appreciated: that there exists ascertainment bias in disease mutation discovery. More generally, our study highlights the parameters that influence the frequencies of deleterious alleles, helping to interpret the relevance of variants of unknown significance based on their allele frequencies, and illuminate some processes that may contribute for the creation and maintenance of mutation load.

Expanded summary*: In this work, we take a first step towards testing long-standing theories about why disease alleles persist in human populations, and in particular whether they simply reflect a balance between mutation and selection, as commonly assumed. Our results highlight the parameters likely to be important in shaping the frequencies of deleterious alleles in natural populations and illuminate the processes that may contribute for the creation and maintenance of mutation load.

We begin our work by reviewing old but often unknown theory about what to expect from mutation and selection in a population of finite size. We then compare expectations from theory (both analytic and simulation-based) to the allele frequencies estimated from 33,370 individuals of European ancestry for highly deleterious mutations. As we show, the predictions fit the data well for highly mutable type of mutations (CpG transitions) but not well for others. Surprisingly, predictions tend to *under*-estimate the frequency at which these disease mutations are observed. In principle, higher than expected frequencies of disease mutations could be due to widespread errors in reporting causal variants, compensation by other mutations, or balancing selection; it is unclear, however, why these factors would affect CpG transitions differently from other mutations. We argue instead that the unexpectedly high frequency of non-CpGti disease mutations likely reflects an ascertainment bias in disease mutation discovery.

In our view, the relevance of our work lies in the finding that mutation-selection-drift balance seems to be a good model for more mutable disease alleles (such as CpG transitions); the broad discussion of factors that are likely to influence the frequencies of deleterious alleles and mutation load, notably the interplay between mutation rate, natural selection and genetic drift; and also the demonstration that biomedical datasets are now starting to be large enough that these sorts of long-standing theories (e.g. mutation-selection balance) can be tested. Moreover, the ascertainment bias in disease mutation discovery that we highlight carries important implications for interpreting the relevance of variants of unknown significance based on their allele frequencies. The results of this work are of potential interest of population geneticists working on theories about mutation load and mutation-selection balance, as well as medical geneticists and evolutionary biologists, among others.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POB-113

Somatic mutation signatures, hotspots and mosaicism in the human lung

Juan Rodriguez-Flores [1,*], Sarah O'Beirne [1], Jacqueline Salit [1], Jason Mezey [1 2], Robert Kaner [1], Ronald Crystal [1]

[1]Genetic Medicine, Weill Cornell Medical College, New York, NY, [2]Biostatistics and Computational Biology, Cornell University, Ithaca, NY, United States

Abstract: While the burden of somatic mutations in cancer tissue and cultured cells is well characterized, studies of somatic mutation burden in live cells obtained from healthy (cancer-free) individuals using ultra-deep whole exome sequencing are scarce. In this study, the somatic mutation burden was quantified lung cells in 99 healthy individuals, diverse with respect to smoking history, ancestry, age and gender. In order to exclude germline mutations, DNA was collected from both blood and lung cells for each individual. Sequencing to over 200* depth across over 50 Mb of target exome was conducted using Illumina 2x100 bp sequencing. Somatic mutations in lung DNA were detected using MUTECT software, strict filtering was applied to the results, and computational prediction of coding variant pathogenicity was assessed using SNPEFF and CADD. The distribution of mutated cell % was quantified across all mutations, and the burden of somatic mutations per megabase was quantified for each individual sample. In addition, an analysis of mutational signatures was conducted. In order to assess the prevalence of somatic mosaicism, sequencing was conducted for two different lung cell types for a subset of 29 individuals, including small airway epithelium (SAE) and alveolar macrophage (AM). These two tissues have different origins, as SAE is derived from stem cells of the basal airway epithelium, while AM cells are derived from bone marrow. The AM cells migrate to the lung, where the environmental exposure is similar to SAE, enabling comparison of somatic mutation burden across these two tissues *in vivo* with consistent environmental factors.

As expected from prior studies, the distribution of somatic mutation frequency followed an exponential distribution, suggesting a minimal influence of positive selection. The burden of somatic mutations for all healthy individuals was observed to be orders-of-magnitude lower than for lung cancer samples assessed using similar methodology. While no significant differences in somatic mutation burden were observed when looking at all samples, a significantly higher burden of somatic mutation was observed among the 10% of SAE samples with the highest somatic mutation burden as compared to the top 10% of AM samples (t-test $p < 0.017$). Mutational signature analysis identified an excess of transversion mutations in SAE (Ts:Tv = 1.95) compared to AM (Ts:Tv = 2.00) (t-test $p < 0.00019$). Somatic mutation hotspots (sites mutated in more than one individual) were also observed in excess in SAE (t-test $p < 0.0004$). Furthermore, an analysis of mosaicism between SAE and AM identified a 1.6-fold excess of somatic mosaic sites in SAE.

This study provides insight into differences in somatic mutation rates in healthy individuals that are specific to cell lineage.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

OW-MM1

Advancing the field of mutational signatures: Insights into mutational mechanisms and clinical applications

Serena Nik-Zainal*

Abstract: A cancer genome is an historical account of the mutagenic activity that has occurred throughout the development of the tumour. Indeed, every mutation matters. While driver mutations were the main focus of cancer research for a long time, passenger *mutational signatures*, the imprints of DNA damage and DNA repair processes that have been operative during tumorigenesis, are also biologically informative[1-5].

We previously outlined the methods for identifying and quantifying base substitution mutational signatures present in primary human cancers[1-5] (http://cancer.sanger.ac.uk/cosmic/signatures). Recently, the intellectual framework of mutational signatures was extended to include six novel rearrangement signatures[6]. Exploring this in >2,500 whole genome sequenced tumours of multiple tumour types, reveals how our early (and rather simplistic) thinking of mutational signatures, requires critical re-evaluation. There are more nuances than previously appreciated.

Diving into the detail of individual mutational signatures[7], we reveal intriguing mechanistic insights into the DNA damage and repair processes that mark the landscape of cancer genomes. In some instances, our findings invite more thoughtful consideration of that relatively binary distinction between drivers and passengers. Furthermore, we demonstrate why our whole genome profiling methods could assist in taking things to a clinical level[6,8], with mutational signatures forming an additional weapon in the arsenal of cancer diagnostics and therapeutic stratification, in the modern war against cancer.

Finally, we find that we can recapitulate cancer mutational signatures using cell-based model systems and arrive at new realisations (INSIGNIA). This powerful approach (humbles the analyst and) emphasises how biological exploration of big data still relies indispensably on experimental work to truly advance understanding.

snz@sanger.ac.uk

@SerenaNikZainal
INSIGNIA project: https://www.mutationsignatures.org
http://cancer.sanger.ac.uk/cosmic/signatures

1. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979-93 (2012).
2. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994-1007 (2012).
3. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* **15**, 585-98 (2014).
4. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-21 (2013).
5. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J. & Stratton, M.R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* **3**, 246-59 (2013).
6. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* (2016).
7. al, G.e. (2017).
8. Davies, H. (2017).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POB-124

Statistically Estimating Somatic Mutation Patterns in Eucalyptus melliodora

Reed A. Cartwright [1,*], Adam J. Orr [1], Rob Lanfear [2]

[1]Arizona State University, Tempe, AZ, United States, [2]Australian National University, Canberra, Australia

Abstract: *De novo* mutations are the ultimate source of genetic variation, and thus are the raw material of evolution. In bilaterian animals, germline tissue is sequestered from somatic tissue and thus mutations that happen in somatic tissue do not enter the germline and contribute to the evolution of species. In plants sequestration does not happen, and somatic mutations can become germline mutations. Here we report the first investigation of somatic mutation patterns in a tree using next-generation sequencing. We have sequenced 24 samples taken from a single *Eucalyptus melliodora* individual from New South Wales; total coverage is over 600x. This individual is especially interesting because it is a somatic mosaic and exhibits two distinct chemical signatures, one of which is resistant to insect herbivory.

To estimate somatic mutation patterns, we have developed a statistical model using hidden-data techniques and Felsenstein's pruning algorithm to calculate the probability of a de novo mutation at a site based on sequencing data. This model not only allows us to estimate somatic mutations, but also to perform phylogenomics, estimating the somatic relationship of our samples. Our estimated somatic phylogeny is consistent with the physical tree topology (a positive control) and allows us to infer the location and patterns of somatic mutation in this individual. An important component of our model is that genotype likelihoods are calculated via mixtures of Dirichlet-multinomial distributions, allowing us to eliminate false positives caused by correlated sequencing error.

Working in *E. melliodora*, we have established that our methods can accurately identify somatic relationships and mutations. This provides credibility for using our methods to study somatic relationships among mammals, for which tissues relationships have not been mapped in detail, but are important for understanding cancer biology and the rates, patterns, and mechanisms for somatic mutations occur in our cells.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POB-112

The role of finescale mutation variation on linked selection in the genome

Rob Ness [1,*]

[1]Biology, University of Toronto @ Mississauga, Mississauga, Canada

Abstract: The rate of *de novo* mutation is known to vary from base to base depending on a number of genomic properties. At very fine-scales, individual bases vary in their mutability and there is also clear evidence for an effect of the local sequence context. At larger scales a number of genome properties influencing de novo mutation rate have been implicated, including transcription level, GC content, or distance from a DNA replication fork. Together these fine- and large-scale drivers of mutability will create heterogeneous evolutionary conditions for genes across the genome. Although mutation is required for adaptive evolution the vast majority of mutations are harmful or effectively neutral. As a result genes subject to higher mutation rates should show elevated divergence at neutral sites, and reduced diversity near selected sites due of background selection and selective sweeps. To test for the influence of mutational heterogeneity on genes we have combined detailed characterization of mutation from experimental lines of *Chlamydomonas reinhardtii* with population genomic analysis. In previous work we used machine learning to predict the mutability of individual nucleotide sites from their genomic properties. Our model allows us to accurately predict the average *de novo* mutation rate for groups of sites. Importantly, the mutability neutral and functionally constrained sites within genes can be predicted separately. Here we test whether variation in mutation rate of genes within a genome alters the rate of molecular divergence. We also investigate whether genes with higher mutation rates show evidence of more harmful or adaptive mutations in the form of reduced nucleotide diversity in and around selected sites.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POB-111

Mitochondrial Mutation Rate, Spectrum and Heteroplasmy in Caenorhabditis elegans Mutation Accumulation Lines of Differing Population Size

Anke Konrad [1,*], Vaishali Katju [1], Ulfar Bergthorsson [1]

[1]Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, United States

Abstract: Understanding population genetic processes in different species contributes towards a deeper understanding of the evolutionary dynamics within and between species. Mutation rates, specifically, provide a first measure of novel genetic variation introduced into genomes that may contribute to divergence between populations and/or species. Hence, genetic markers within the mitochondrial genomes of metazoans, in large part due to their increased mutation rates, have long been employed to investigate the degree of relatedness within and between species and populations. In order to better understand the dynamics of spontaneous mutational processes affecting mitochondrial genomes in small populations under varying degrees of selection, we evolved a total of 35 mutation accumulation lines of *Caenorhabditis elegans* hermaphrodites with differing constant population sizes (N = 1, 10, and 100) for up to 409 generations. This experimental design allows for the isolation of the spontaneous mutation rate estimate in the near-absence of selection given by the N = 1 lines. Analyzing the larger populations provides a better picture of the retention of mutations under increasing degrees of selection. Novel mutations within 85 individual genomes across the 35 MA lines were identified using Illumina paired-end sequencing and compared to the mitochondrial genome of their ancestor. The analysis led to an estimated spontaneous mutation rate of $1.05 \times 10{-}7$ mutations/site/generation with a strong G/C→A/T base substitution bias. The same substitution bias was found when the analysis was conducted on 38 natural isolates of *C. elegans*, which suggests that the observed low synonymous site G+C content in *C. elegans* can be attributed to the mutational process, rather than subsequent selection, as has previously been suggested. Despite previous reports that heteroplasmy is rare in mitochondria of *C. elegans*, the vast majority of mutations detected in this study were heteroplasmic, with very few reaching frequencies close to fixation by the end of the experiment. Both, frameshift and nonsynonymous mutations showed a negative correlation to increasing population size, implying that increasing degrees of selection in the larger populations effectively eradicate these deleterious mutations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POB-116

Microsatellite and transcriptional changes in a malaria mutation accumulation experiment

Marina McDew-White [1,*], Standwell Nkhoma [1], Tim J. C. Anderson, Ian Cheeseman [1]

[1]Genetics, Texas Biomedical Research Institute, San Antonio, United States

Abstract: Microsatellite sequences are widely assumed to evolve neutrally, but can also play an important role in bacterial pathogenesis and human disease and have been proposed as a means for fine-tuning transcription, so the neutral assumption maybe questionable. The malaria parasite *Plasmodium falciparum* is absurdly AT-rich and contains microsatellites on average every ~1 kb across the 23 Mb genome. This project was designed to determine the microsatellite mutation rate in malaria parasites, and to investigate whether microsatellites are key determinants of transcriptional change in his pathogen. We maintained 40 parasite lines derived from a single parasite cell for 83 to 261 days, with frequent bottlenecking to a single cell to minimize effective population size, allowing us to measure mutations accumulated over >15000 mitoses. We illumina sequenced genomes of both progenitor and end-point mutation accumulation (MA) parasite lines in duplicate and called microsatellites using a validated LOBSTR pipeline. We scored 21,654 microsatellite loci, with 13937 scored in >50% of MA lines. Calls were 99.47% concordant in duplicate sequence runs from independent sequence libraries. We observed 1353 microsatellite mutations, giving rates of $1.04 \times 10^{5} – 3 \times 10^{-7}$ /cell division for different motif lengths: hence in a single infection (10^{11} parasites) we expect to see 10^{4} - 10^{6} independent mutations at any single microsatellite locus. Furthermore, many of these microsatellites are found either within or close to gene sequences. We are currently examining transcript variation across the parasite lifecycle in a subset of these MA parasite lines to determine how microsatellite length changes influence transcript abundance.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POB-430

TRANSLATIONAL EFFICIENCY AND THE EVOLUTION OF POSITION-DEPENDENT CODON USAGE

Nelson Morrow [1,*], Ashley Teufel [2], Alon Diament [3], Tamir Tuller [4], C O. Wilke [2]

[1]Department of Physics, [2]Department of Integrative Biology, The University of Texas, Austin, United States, [3]Department of Biomedical Engineering, Tel Aviv University, Tle Aviv, [4]Department of Biomedical Engineering, Tel Aviv University, Tel Aviv, Israel

Poster: The translation of mRNA by ribosomes is the fundamental process underlying protein synthesis. It has been observed that mRNA sequences with codons corresponding to abundant tRNAs are more highly expressed, implying a relationship between expression level and codon usage. This observation has implications for position-dependent codon usage. In fact, the first 30-50 codons have been shown to translate with relatively low efficiency, on average. This initial group of low efficiency codons is often referred to as a ramp. It is hypothesized that these ramps serve to restrict the number of ribosomes allocated to any one sequence at a time to prevent ribosomal traffic jams. Here we simulate the codon evolution of a network of yeast genes evolving under selection for translation efficiency and minimizing translation error with the use of a ribosome flow model (RFM), to examine if these constraints result in formation of ramps. This ribosome flow model considers the dynamic nature of the translation process, modeling translation rates, protein abundance levels, and ribosomal densities interconnected through a pool of free ribosomes. We expect that under some parameterizations of our RFM selection for translation efficiency and the minimization of translation error will result in ramps at the beginning of some sequences in order to discourage ribosomal traffic jams and increase overall network efficiency.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POA-421

TRACING EVOLUTIONARY TRAJECTORIES THROUGH POPULATION DUPLEX SEQUENCING

Han Mei [1,*], Anton Nekrutenko [1]

[1]The Pennsylvania State University, University Park, United States

Poster: In bacterial experimental evolution studies genomic changes are often determined via analysis of several timepoints - aliquots that are drawn from the experimental populations, amplified, and subjected to DNA sequencing. While illustrative, such sampling can be viewed as an extreme case of bottleneck in which only genetic changes at high frequency are revealed but no information about the underlying mutational dynamics is retained. Here we present results of a turbidostat experiment in which at every timepoint approximately ⅔ of the evolving population is sampled and, without amplification, subjected to duplex sequencing designed to reveal genetic changes at very low frequencies. The population consists of *Escherichia coli* DH5▫▫ cells transformed with pBR322 plasmid and grown in a tetracycline-containing media. The high sensitivity of our mutation detection strategy allows us to trace substitutions as they occur through time within pBR322, which is being evolutionarily "optimized" towards higher expression of *tetR* gene required for tetracycline resistance.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

OW-MM2

Generating hypotheses about mutational mechanisms from large-scale data

Shamil Sunyaev [1,*]

[1]United States

Abstract: Biochemistry and model system experimental genetics have accumulated massive knowledge on possible mechanisms of mutation. However, relative contributions of replication infidelity and mis-repaired DNA damage to real mutagenesis remain unknown. We clearly lack mechanistic hypotheses on specific biological processes that dominate mutagenesis in populations rather than in the experimental setting.Sequencing datasets on parent-child trios provide a source of data on de novo germ-line mutations. Although data on true de novo mutations are still sparse, they can be efficiently combined with large-scale data on rare population genetic variation. The combined analysis of germ-line de novo mutations, population sequencing data and cancer genomics data is highly informative about statistical properties of newly arising mutations. Combining these data with epigenomic features such as replication timing, gene expression and chromatin accessibility is helpful in generating mechanistic hypotheses. Studying the effect of the tumor genotype on mutational patterns provides an additional instrument for the analysis of forces underlying mutagenesis.We analyzed germ-line and somatic mutation data with respect to the direction of replication and transcription. The observed patterns can be interpreted in light of experimental data on DNA repair efficiency. Collectively, the data shed light on the mechanism maintaining strand bias of mutations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

OW-MM6

Polymerization kinetics deciphered using PacBio sequencing: Non-B DNA affects polymerase progression and error rate

Kateryna Makova [1,*], Wilfried Guiblet [1], Marzia Cremona [1], Monika Cechova [1], Robert Harris [1], Iva Kejnovska [2], Kristin Eckert [3], Eduard Kejnovsky [2], Francesca Chiaromonte [1]

[1]Penn State University, University Park, United States, [2]Institute of Biophysics, Brno, Czech Republic, [3]Penn State University, Hershey, United States

Abstract: Studies of individual loci demonstrated that non-B DNA (e.g., G-quadruplexes, Z-DNA, cruciforms, and slipped structures) causes polymerase stalling and replication errors, leading to genome instability. To date, these important effects of non-B DNA have not been investigated on a genome-wide scale. Here we explore DNA polymerization kinetics using human whole-genome data resequenced with Pacific Biosciences (PacBio) technology. In addition to base calling, this technology registers the time between incorporation of two consecutive bases, or InterPulse Duration (IPD). Using novel Functional Data Analysis techniques, we demonstrate that non-B DNA motifs lead to polymerization kinetics that is significantly deviant from that observed outside of such motifs. These alterations depend on motif identity and DNA strand involved. For instance, G-quadruplexes display a strong strand-specific deceleration in polymerization. We confirm this computational result with an experimental analysis of circular dichroism and show that polymerization kinetics at these motifs depends on their thermostability. Among other notable effects of non-B DNA structures are a significant polymerization acceleration at AT-rich homopolymer motifs and a periodic change in polymerization speed at microsatellite motifs. Importantly, several non-B motifs significantly increase the rate of PacBio sequencing errors. However, this effect explains only a small proportion of variation in polymerization speed, suggesting that most of the variation is biological. Furthermore, base composition and epigenetic modifications at non-B DNA motifs cannot explain the observed variation in polymerization dynamics. Finally, we extend our *in vitro* observations to *in vivo* effects via studying the relationship between non-B DNA motifs and human genetic diversity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POB-114

High-quality mitochondrial DNA sequencing of single oocytes

Barbara Arbeithuber [1,*], Nicholas Stoler [2], Anton Nekrutenko [2], Kateryna Makova [1]

[1]Department of Biology, [2]Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, United States

Abstract: With the development of next generation sequencing (NGS) about a decade ago, the milestone was set to obtain DNA sequences from a single cell. By combining whole genome amplification (WGA) with NGS, the nuclear, as well as the mitochondrial, genome of single cells could be sequenced for a wide range of different cell types, including oocytes.
Despite these great advances in single-cell sequencing, the analysis of rare variants in the mitochondrial genome (mtDNA) is hindered by the high error rates associated with the sequencing methods, as well as false positive mutation calls resulting from DNA lesions. Duplex sequencing, a specialized sequencing method developed by Schmitt and colleagues (first published in 2012) has the power to greatly reduce such errors (down to the levels below 10^{-7}) by independently tagging each of the two strands of a DNA duplex, followed by the amplification and sequencing of both strands separately. While true mutations are found at both DNA strands, false positive mutations can be identified since they are only present at one of the strands or in several reads of one strand.
While a high amount of DNA is required for the original duplex sequencing protocol, we were able to improve the library preparation efficiency and successfully perform duplex sequencing on single murine oocytes. In addition to improvements in library preparation, the implementation of a more efficient bioinformatic pipeline for single-strand and duplex consensus formation – *Du Novo* – allows reliable detection of mtDNA mutation frequencies below 0.3%.

Expanded summary*: Mutations in mitochondrial DNA (mtDNA) contribute to a variety of diseases. Most of the disease-causing mtDNA mutations are associated with heteroplasmic sites, i.e. mtDNA sites for which multiple variants coexist within an individual [1]. When the frequency of a disease-associated variant exceeds a threshold, symptoms occur [2]. While only ~1 in 4300 humans exhibits symptoms of a mtDNA-associated disease, it was shown that 1 in 200 infants harbors a known pathogenic mtDNA mutation [3]. However, in this study only 10 most common pathogenic mtDNA mutations at frequencies >1% were analyzed, leading to the following questions: What is the real frequency of (potentially pathogenic) mutations in mtDNA. What is the frequency of their carriers? When do the mtDNA mutations arise during human ontogenetic development? How are they transmitted to the next generation (mtDNA is maternally transmitted)? And how can the level of heteroplasmy increase from one initial mutation to pathogenic levels?
To address these questions, it is necessary to measure and quantify low-level mtDNA variants with high accuracy and throughput. Furthermore, the study of mtDNA in single oocytes will provide important information on processes such as the emergence and inheritance of low-frequency heteroplasmies. Therefore, the development of a method to obtain high-quality mtDNA sequences from single oocytes combined with the analysis of somatic tissue samples will allow us to study in detail the inheritance of low-level heteroplasmic sites and estimate the germline mtDNA mutation rate.
Technologies that provide the required high throughput and single-molecule sensitivity include next generation sequencing (NGS), often used in combination with whole genome amplification (WGA) for single-cell analysis. However, given the high error rates of these technologies and false positive mutation calls resulting from DNA lesions, they are unsuitable for detecting rare variants. Schmitt and colleagues developed a method termed “duplex sequencing”, which greatly reduces errors resulting from DNA damage and amplification in NGS experiments by independently tagging each of the two strands of a DNA duplex, followed by the amplification and sequencing of both strands separately [4]. While true mutations are found in both DNA stands, artifacts resulting from DNA damage, sequencing, and PCR can be identified since they are only present in one of the strands or in several reads of one strand.
Although in the published duplex sequencing protocol the minimum DNA amount for library preparation is given with 100 ng [5], we were able to apply duplex sequencing to single murine oocytes (containing a few picograms of total DNA). This could be established on the one hand by improving the library preparation efficiency by minimizing sample purification steps and improving ligation efficiency, but also by the implementation of a more efficient bioinformatic pipeline for single-strand and duplex consensus formation:

Du Novo [6]. With these improvements, we were able to reliably detect mtDNA mutation frequencies below 0.3% in single oocytes, demonstrating the potential of the newly developed protocol to accurately measure low-level heteroplasmies and germline mutations in single oocytes.

References:

1. Pereira, L., Freitas, F., Fernandes, V., Pereira, J.B., Costa, M.D., Costa, S., Maximo, V., Macaulay, V., Rocha, R. and Samuels, D.C. *The diversity present in 5140 human mitochondrial genomes.* Am J Hum Genet. 2009. 84(5): p. 628-40.
2. Wallace, D.C. and Chalkia, D. *Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease.* Cold Spring Harbor perspectives in biology. 2013. 5(11): p. a021220.
3. Elliott, H.R., Samuels, D.C., Eden, J.A., Relton, C.L. and Chinnery, P.F. *Pathogenic mitochondrial DNA mutations are common in the general population.* Am J Hum Genet. 2008. 83(2): p. 254-60.
4. Schmitt, M.W., Kennedy, S.R., Salk, J.J., Fox, E.J., Hiatt, J.B. and Loeb, L.A. *Detection of ultra-rare mutations by next-generation sequencing.* Proc Natl Acad Sci U S A. 2012. 109(36): p. 14508-13.
5. Kennedy, S.R., Schmitt, M.W., Fox, E.J., Kohrn, B.F., Salk, J.J., Ahn, E.H., Prindle, M.J., Kuong, K.J., Shen, J.C., Risques, R.A. and Loeb, L.A. *Detecting ultralow-frequency mutations by Duplex Sequencing.* Nat Protoc. 2014. 9(11): p. 2586-2606.
6. Stoler, N., Arbeithuber, B., Guiblet, W., Makova, K.D. and Nekrutenko, A. *Streamlined analysis of duplex sequencing data with Du Novo.* Genome Biol. 2016. 17(1): p. 180.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

OW-MM3

DNA editing of LTRs by APOBECs accelerate vertebrates genome evolution

Binyamin Knisbacher [1], Erez Levanon [1,*]

[1]Bar-Ilan University, Ramat Gan, Israel

Abstract: Long terminal repeat retrotransposons (LTR) are widespread in vertebrates and their dynamism facilitates genome evolution. However, these endogenous retroviruses (ERVs) must be restricted to maintain genomic stability. The APOBECs, a protein family that can edit C-to-U in DNA, do so by interfering with reverse transcription and hypermutating retrotransposon DNA. In some cases, a retrotransposon may integrate into the genome despite being hypermutated. Such an event introduces a unique sequence into the genome, increasing retrotransposon diversity and the probability of developing new function at the locus of insertion. The prevalence of this phenomenon and its effects on vertebrate genomes are still unclear. In this study, we screened ERV sequences in the genomes of over 100 diverse species and identified hundreds of thousands of edited sites in multiple vertebrate lineages, including placental mammals, marsupials, and birds. Numerous edited ERVs carry high mutation loads, some with greater than 350 edited sites, profoundly damaging their open-reading frames. For many of the species studied, this is the first evidence that APOBECs are active players in their innate immune system. Unexpectedly, some birds and especially zebra finch and medium ground-finch (one of Darwin's finches) are exceptionally enriched in DNA editing. We demonstrate that edited retrotransposons may be preferentially retained in active genomic regions, as reflected from their enrichment in genes, exons, promoters, and transcription start sites, thereby raising the probability of their exaptation for novel function. In conclusion, DNA editing of retrotransposons by APOBECs has a substantial role in vertebrate innate immunity and may boost genome evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

OW-MM5

Spatial chromatin organization and distribution of somatic mutations in cancer genomes

Kadir C. Akdemir*

Abstract: The hierarchical folding of vertebrate DNA is closely associated with transcriptional regulation. Recent chromosome conformation studies suggest that mammalian chromosomes are structured into tissue-invariant topologically associating domains (TADs) where the regions within a domain interacts more frequently together than with regions in other domains. Genes within the same TADs represent similar expression and histone-modification profiles. Here, to understand the distributions of somatic mutations in human cancers, we utilized data from 2750 high-coverage whole genome sequences across 45 different cancers with paired normal samples. Our analysis revealed a strong correlation between the mutational distributions in human cancers and the spatial organization of the genome. Transcriptionally active TADs contain less mutation burden compared to inactive TADs, as a result regional mutation rates are drastically different around the boundaries delineating epigenetically distinct domains. However, somatic mutation distributions show variations based on the underlying exogenous mutagens. For example, UV-mediated mutations are enriched in epigenetically inactive domains of melanoma genomes compared to the mutations caused by a plant-based carcinogen, aristolochic acid (AA) in urinary tract cancer genomes. Taken together, our analyses reveal new insights about genome architecture and mutational distributions in human cancers.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POB-117

High rate of mutation in satellite DNA, but does selection see it?

Jullien Flynn [1,*], Ian Caldas [2], Melania Cristescu [3], Andrew Clark [1]

[1]Molecular Biology and Genetics, [2]Biological Statistics and Computational Biology, Cornell University, Ithaca, United States, [3]Biology, McGill University, Montreal, Canada

Abstract: Eukaryotic genomes contain large amounts of tandemly-repeated satellite DNA, but its composition varies greatly among even closely related species. The mechanisms and rates of mutation in satellite DNA remain unknown, and whether selection influences satellite DNA evolution is debated. We used whole-genome sequences of 28 mutation accumulation (MA) lines of *Daphnia pulex* in addition to six isolates from a population originating from the same progenitor to both estimate mutation rates of satellite sequences and evaluate the selection regime acting upon them. We found that mutation rates of individual satellite sequence "kmers" were very high and linearly related to the kmer's abundance, ranging from additions/deletions of 0.3 – 105 copies per generation. The population isolates showed a strong signal of purifying or stabilizing selection with a lower mutation rate and reduced variation in abundance across almost all kmers. We also found that new repeat sequences are often generated from pre-existing ones. We conclude that satellite DNA evolves so rapidly largely as a consequence of its high mutation rate, but at the same time, its evolution is demonstrably constrained by selection.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POB-118

Specificity of the DNA mismatch repair system (MMR) and mutagenesis bias in bacteria

Hongan Long [1,*], Samuel Miller [1], Emily Williams [1], Michael Lynch [1]

[1]Biology, Indiana University, Bloomington, United States

Abstract: The mutation rate of an organism is influenced by the interaction of evolutionary forces such as natural selection and genetic drift. However, the mutational spectrum (i.e. the distribution of different types of mutations) is heavily influenced by various intracellular processes such as oxidation, deamination, DNA replication repair. Using original and recently published whole-genome sequences of mutation accumulation lines for Bacillus subtilis subsp. subtilis, Deinococcus radiodurans, Escherichia coli and Pseudomonas fluorescens, we quantify the general or distinct features of mutagenesis, and the efficiency and specificity of MMR (DNA mismatch repair) in bacteria. We do find some general MMR patterns in all four bacteria: repair efficiency is heavily influenced by the neighboring base composition, and only recognizes indels < 4 bp in length. Unlike all other bacteria in this study, D. radiodurans has an insertion bias, an MMR system that preferably repairs deletions, and a very low MMR repair efficiency.

Expanded summary*: DNA repair is a collection of multiple pathways to ensure genome stability and efficiently fixes DNA damages, which could reduce fitness or even lead to death of cells. Previous studies have shown that the process of DNA repair is biased and not random. However, these biases are never rigorously quantified, because of the numerous assumptions and the limited genomic regions of the reporter constructs that are used in fluctuation tests—the method for estimating mutation rate in most earlier mutational studies. By comparing mutation patterns of MMR− and MMR+ strains of multiple bacteria species using the mutation accumulation combined with whole-genome sequencing method, the efficiency and specificity of the MMR repair systems can be accurately quantified, as well as the MMR efficiency variation and determinants.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POB-110

Highly accelerated rates of heritable large-scale mutations under heavy metal exposure

Frederic Chain [1,2,*], Jullien Flynn [3], James Bull [2], Melania Cristescu [2]

[1]UMass Lowell, Lowell, United States, [2]McGill University, Montreal, Canada, [3]Cornell University, Ithaca, United States

Abstract: Mutation rate variation across taxa has been under intense investigation during the last decades. Despite the interest in understanding the evolutionary and ecological factors influencing mutation rates, little is known about the extent to which environmental stressors accelerate mutation rates and the genetic load of populations. Moreover, most studies on mutation rates focus on point mutations rather than large-scale deletions and duplications with mutational mechanisms possibly more readily induced by stress. We estimated mutation rates in *Daphnia pulex* exposed to environmental stressors and quantified the effect of selection on copy number variations (CNVs). We conducted a mutation accumulation (MA) experiment in which selection was minimized, coupled with an experiment in which a population was propagated under competitive conditions in a benign environment. After around 100 generations of propagation, we sequenced 60 genomes and found significantly higher rates of deletions and duplications in MA lines exposed to ecologically relevant concentrations of heavy metals compared to control lines. Whereas control MA lines had gene CNV rates comparable to other multicellular eukaryotes (1.1×10^{-6} per gene per generation), copper induced 1.3x higher rates and mixtures of nickel and copper increased rates by 4.4x. Non-MA isolates that experienced selection carried 10x fewer CNVs and a sixth of the proportion of CNVs overlapping genes compared to control MA lines, providing evidence that CNVs contribute to mutational load. Our CNV breakpoint analysis revealed that nonhomologous recombination associated with regions of DNA fragility is the primary source of CNVs, plausibly linking metal-induced DNA strand breaks with higher CNV rates.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POB-122

The de novo mutation rate of transposable elements in the Chlamydomonas reinhardtii genome

Fathiya Mohamed [1,*], Rob Ness [1 2]

[1]Department of Biology, University of Toronto Mississauga, Mississauga, [2]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada

Abstract: Transposable elements (TEs) are a class of repetitive DNA sequences that can move positions in the genome. TEs have an important role in driving key aspects of evolution including genome size as well as the generation of novel phenotypic variation. Despite their high frequency in many eukaryote genomes, they are relatively difficult to characterize in genomic data because of their repetitive nature. Studies measuring the rate of de novo TE insertions have traditionally relied on phenotypic or PCR markers and were not able to directly identify the location of these new insertions. However, new algorithms that exploit the patterns of next generation sequencing alignments can detect the presence of new TEs not present in the reference genome. In our study, we used mutation accumulation (MA) to analyze the rate and distribution of de novo TE insertions in the sequenced genomes of 85 *Chlamydomonas reinhardtii* lines from 6 genetically distinct strains. In combination with high-throughput sequencing, MA studies allow for characterizing the properties of spontaneous mutations in the genome under minimal selection. We have developed a pipeline that combines available TE calling software with de novo assemblies of putative TE insertions to accurately identify new TEs. We will present data that measure the rate of spontaneous mutation in the *Chlamydomonas reinhardtii* genome, explore the presence of transposition hotspots as well as compare differences in TE distributions between MA lines and natural populations.

Statement: I am in my final year of study at the University of Toronto working towards the completion of a BSc in Biology. The most rewarding aspect of my undergraduate degree was being able to conduct multiple independent research projects over the past three years. This year, I have been drawn to work on evolutionary genomics which has become the focus of my thesis project. The field of evolutionary genetics in general has inspired me to pursue further research as a graduate student. SMBE brings together many leading researchers in molecular biology and evolution at this meeting, and I hope that being exposed to their new ideas and techniques will influence the course of my research career.
The greatest benefit of attending this meeting as an undergraduate researcher is the mentorship SMBE facilitates to enrich the minds of aspiring academics. As a female belonging to an underrepresented ethnic minority, this mentorship experience paired with the chance to present my work provides a unique opportunity to enhance my research background.
As someone who hopes to go into academia, it would be valuable to get the chance to interact with those who are established in the field and learn more about what a career in research entails through this process. Although I am early in my research career, I have devoted over a year of work towards characterizing the rate of de novo TE mutations and this project will contribute both to our understanding of genome evolution and to the SMBE 2017 meeting.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

POB-121

Precise estimation of genome-wide mutation rate and spectrum in the halophilic archeon Haloferax volcanii

Sibel Kucukyildirim [1 2,*], Megan Behringer [1], Emily Williams [1], Thomas Doak [1], Michael Lynch [1]

[1]Biology, Indiana University, Bloomington, United States, [2]Biology , Hacettepe University, Ankara, Turkey

Abstract: Background: Spontaneous mutations play a central role in all evolutionary processes, and like most phenotypic characteristics, the rate of mutations is determined by an interaction of environmental and genetic factors. Investigating organisms adapted to life in extreme habitats may further our understanding of the mechanisms of genetic stability. In this study, we report the genome-wide measurement of spontaneous mutations in the halophilic archeon *Haloferax volcanii* using a direct and unbiased method: mutation-accumulation experiments and deep whole-genome sequencing. *H. volcanii* is a key model organism, not only for the study of halophilicity, but also for archaeal biology in general. This work enables us to explore evolutionary forces and molecular mechanisms that may have shaped the mutation rate and spectrum of *H. volcanii.*

Results and Discussion: Our methods measure genome-wide rate, spectrum and distribution of spontaneous mutations. Whole-genome sequencing of 54 mutation accumulation lines of *H. volcanii* after an average of ~3000 cell divisions yielded a base-substitution mutation rate of 0.0012 per genome per generation, which is surprisingly similar to the consensus value of mesophilic organisms with DNA genomes. Transitions were found more frequently than transversions, and contrary to all other characterized G/C-rich prokaryotes, in *H. volcanii* an A/T mutation bias is observed. We do not observe the short insertion/deletion bias observed in previous reporter-gene experiments with *H. volcanii.* Large-scale deletions are restricted to secondary chromosomes/plasmids. Whole genome sequencing of mutation accumulation lines provides the comprehensive insights of mutations and reveals how the mutational process affects genomic architecture.

Expanded summary*: Spontaneous mutations play a central role in all evolutionary processes, and like most phenotypic characteristics, the rate of mutations is determined by an interaction of environmental and genetic factors. My research focuses on the general question of how mutations accumulate in genomes. There has been a great deal of theoretical and experimental work (much of it from the Lynch lab) on what the intrinsic mutation rate of organisms is expected to be, and what it actually is. This rate is essentially a result of how often the replicative DNA polymerase makes mistakes, and the cell's repair pathways fail to fix these mistakes. A primary determinant is how big the genome is, resulting in organisms with very different sized genomes having the same absolute number of mutations per generation—implying that their polymerases must have different error rates per base pair.

While there have been many mutation accumulation (MA) assays done with bacteria and metazoa, very few have been done with archaea—and these have used indirect methods. My work in the Lynch lab has been on mutation accumulation experiments in diverse organisms, to broaden the current knowledge on rates and molecular spectra of spontaneous mutations. I have completed analyses for *Haloferax volcanii*, which I plan to present at this meeting, and will continue working on MA lines of diverse organisms.

H. volcanii offers an excellent model system, not only for the study of halophilicity, but also for archaeal biology in general. Although *H. volcanii* is an obligate halophile requiring high salt conditions to grow, because it is aerobic and mesophilic, *H. volcanii* can be grown in conditions much like those used for other model organisms, such as the bacterium *Escherichia coli* and the model yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*.

In this study, we investigate how the mutation rate and spectrum has evolved in organisms adapted to life in extreme habitats. For the first time, we report the genomic mutation rate and spectrum for the *H. volcanii*, derived using the mutation accumulation method followed by whole-genome sequencing of replicate lines. The base-substitution mutation rate is estimated to be 0.0012 per genome per generation, which is surprisingly similar to the consensus value of mesophilic organisms. However, mutations in *H. volcanii* A/T biased, opposite to what is observed in most studied bacteria with G/C-rich genomes. To our knowledge, our study provides the first precise measurement of the genetic fidelity maintained by an Achaea and provides a more complete view of how several mechanisms of mutation, mutation repair, and bias act simultaneously to produce the raw material for evolution across the tree of life.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Mutational mechanisms

OW-MM4

Clustered de novo mutations with large intra-mutational distance contribute to the maternal age effect

Wendy Wong [1,*], Jakob Goldmann [2], Vladimir Seplyarskiy [3], Thierry Vilboux [1], Dale Bodian [1], Benjamin Solomon [4], John Deeken [1], Christian Gilissen [5], John Niederhuber [1 6]

[1]Inova Translational Medicine Institute (ITMI), Inova Health Systems, Falls Church, [2]Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, [3]Division of Genetics, Brigham and Women's Hospital, Boston, [4]GeneDx, Gaithersburg, United States, [5]Department of Human Genetics, Donders Centre for Neuroscience, Radboud University Medical Center, Nijmegen, Netherlands, [6]Johns Hopkins University School of Medicine, Baltimore, United States

Abstract: Clustered mutations have been found in both somatic mutations in cancer genomes, and in inherited de novo mutations (DNMs) in the germline. Evidence suggests that these clustered mutations are derived from a single mutational event. While various mutational mechanisms of clustered mutagenesis have been discovered in tumorigenesis, they are largely elusive in gametogenesis. In this study, we identified 1,796 clustered DNMs (cDNMs) with inter-mutation distances less than 20kb in the whole-genome sequencing data from 1,291 parent-offspring trios. Using the familial and sequence reads information we successfully determined the parent-of-origin of 660 cDNMs, of which 49 percent are paternal and 51 percent are maternal. We confirmed the significantly higher proportion of C>G substitutions in clusters that has been observed in previous studies. We discovered that regions with the high fraction of clustered C>G in the 1000 genomes data are highly correlated with regions enriched for maternal DNMs and that maternal cDNMs are enriched by coordinated C>G/C>G clusters. Finally, we found that specifically cDNMs with large intra-mutational distance (>1kb) contribute to the maternal age effect, and not to the paternal age effect. Only these large clusters are positively correlated with maternal recombination rates, suggestive of mutational mechanisms related to double strand breaks. Taken together, our study sheds light on the distinct mutagenic mechanisms in clustered DNMs and reveals their significant contribution to the maternal age effect.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS6

Rewiring key gene interactions during embryonic development alters larval morphology in the sea urchin genus Heliocidaris

Gregory Wray [1,*], Lingyu Wang [1], Allison Edgar [1]

[1]Biology, Duke University, Durham, United States

Abstract: Morphological traits are the result of numerous gene interactions that unfold over the course of development, yet the ways that mutations reconfigure these interactions to produce adaptive changes remains poorly understood. The developmental gene regulatory network (GRN) that produces the larva of sea urchins has been reconstructed in considerable detail. We are using this information to identify specific changes in gene interactions that are responsible for a dramatic change in larval morphology between two closely related species in the genus *Heliocidaris*. Based on our earlier evolutionary analyses of developmental transcriptomic time-series, we are now experimentally manipulating the expression of individual genes within the GRN using knock-downs and mis-expression. As expected, we find that some gene interactions are likely conserved in the species with the derived morphology, although some of these interactions now occur at different relative times during development. More significantly, we also find that several key gene interactions are qualitatively altered. These include a crucial sub-circuit of the GRN that specifies a set of cells that form the earliest signaling center in embryo, which initiates a cascade of subsequent cell fate decisions. Previous studies have shown that key interactions within this GRN have been conserved for at least 250 my, prompting the hypothesis that these gene interactions are so integral to early development that they are no longer evolvable. Our finding that some of these same gene interactions changed during the ~2 my since the two *Heliocidaris* species diverged suggests that stabilizing selection, rather than developmental constraint, is responsible for these highly conserved gene interactions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OT-OS1

Family quarrels in seeds result in molecular arms races

David Queller [1,*], Katherine Geist [1], Joan Strassmann [1]

[1]Department of Biology, Washington University in St. Louis, St. Louis, United States

Abstract: Kin selection theory predicts that parents and offspring should be in conflict over the amount of care parental care provided. Conflict, in turn, is predicted to drive evolutionary arms races. Kin selection is little studied in plants, but flowering plants offer a particularly interesting case because the triploid endosperm takes over most of the role of nutrient acquisition from the mother. We tested and confirmed multiple predictions about how kin selected conflict should increase rates of adaptive molecular evolution in seeds of *Arabidopsis*. There is (1) more adaptive evolution in genes expressed in seeds than in other specialized organs, (2) more in endosperms and maternal tissue than in embryos, and (3) more in the specific compartments involved in the transfer of nutrients than in those that do not. These results show that kin selection is important in plants, and that parent-offspring conflict can drive evolutionary arms races despite the conflict-reducing role of kinship.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-354

Expression of endogenous retroviruses in canine cancer cell lines.

Abigail Jarosz [1,*], Malika Day [1], Julia Halo-Wildschutte [1]

[1]Biology, Bowling Green State University, Bowling Green, United States

Abstract:

In the canine reference genome, there is only a 0.15% presence of endogenous retroviruses (ERVs), known as CfERVs. This is a relatively low percentage considering ERVs make up 8% of the human genome. Of the ERVs found in canids, most are believed to be from considerably old infections of retroviruses. Due to the old age of these insertions, there are a multitude of accumulated mutations that prevent the provirus from coding for functional genes. With low representation in the canine genome as well as the apparent lack of present exogenous retroviruses, it has been previously assumed that the integration of ERVs in canines was extremely rare or nonexistent. Despite this, there have been reports of retroviral activity in tissues taken from canine tumors and cancer cell lines. In recent findings, young copies of ERVs, known as CfERVFc1, have been identified that have sequence similarity to the mammalian ERV-Fc/W groups in the boxer reference. Some of these proviruses had either completely intact or nearly intact open reading frames. Upon further investigation of these proviruses, the LTRs were shown to have a low amount of mutations, suggesting infection within at least the last ~0.48 million years. Additionally, the envelope gene present in the recently found proviruses show to be specific to the canine species again suggesting a more recent infection of XRV than what has been theorized.

To our knowledge, there are no current infectious retroviruses that have been found in any canines or wild canid. Curiously, there have been findings of reverse transcriptase activity along with gamma-type C particles in tumor tissues. We hypothesize that this 'young' CfERV lineage is transcriptionally active in canine cancers. Knowing of the presence of these more recently active endogenous retroviruses, we wanted to investigate if there is expression present in cancer cell lines. To begin the search of the expression of single provirus loci in canine cancer cell lines, we selected and cultured four canine cancer cell lines. Once all the cells were grown, we preformed RNA extractions and then reverse transcribed the RNA into cDNA. By analyzing cDNA from extracted RNA, we were able to confirm not only the presence of the CfERV integrants in the genome of those samples, but we also confirmed that, under certain conditions, the integrants are transcriptionally active. Primers were designed to amplify multiple, highly conserved regions of the *pol* gene. To confirm the expression from PCR was in fact the *pol* sequence, the PCR products were transformed in competent cells. Products of these transformed cells were sequenced at individual loci. We also plan to analyze the distribution of these individual loci in cancer and disease to help us better understand the biological implications of endogenous retroviruses in the canine health, a close medical model to humans.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OT-OS3

The advent of agriculture shaped innate immune responses to pathogens in humans

Genelle Francis Harrison [1,*], Joaquin Sanz-Remon [2], Christina Bergey [3], Anne Dumaine [4], Vania Yotova [4], Jean-Christophe Grenier [4], Lluis Quintana-Murci [5 6 7], George Perry [3], Luis Barreiro [4 8]

[1]Human Genetics, McGill University, [2]Department of Biochemistry, Universite´ de Montreal, Montreal, Canada, [3]Departments of Anthropology and Biology, Pennsylvania State University, University Park, United States, [4]Department of Genetics, CHU Sainte-Justine Research Center, Montreal, Canada, [5]Unité de Génétique Évolutive humaine, Institut Pasteur, [6]Centre national de la recherche scientifique, [7]Université Pierre et Marie Curie, Paris, France, [8]Department of Pediatrics, Universite´ de Montreal, Montreal, Canada

Abstract: The shift of societies from a hunter-gatherer to agricultural method of subsistence in Africa is considered to have facilitated the emergence of many pathogens. Yet, the extent to which hunter-gatherers and agricultural populations diverge in their immune response has not been evaluated, nor have we gauged the role of selection in contributing to these differences. Here, we collected peripheral blood mononuclear cells (PBMCs) from both hunter-gatherer (Batwa) and agricultural (Bakiga) populations in Uganda. We stimulated the PBMCs using viral (Gardiquimod–GARD) and bacterial (lippopolysacharide–LPS) ligands and looked for a divergence in transcriptional regulation of innate immune response. We evaluated transcriptional differences between the Batwa and Bakiga populations (PopDE), and found 1,664 PopDE genes that differed in their overall expression for LPS and 2,242 PopDE for GARD. Among these PopDE genes we found an increase in the anti-viral activity in the Batwa population with the increased expression of genes in interferon (IFN) pathways. We next mapped expression quantitative trait loci (eQTL) for 1,097 genes. Genes with cis-eQTL are enriched among PopDE genes, suggesting that a significant fraction of transcriptional differences are genetically controlled. However, differences in IFN-signaling cannot be explained by cis genetic regulatory variants suggesting that either a trans-eQTL or non-genetic factors are responsible for the increased activity of anti-viral responses in the Batwa. Several of the eQTL driving population differences in immune regulation were targeted by positive selection, reinforcing the long-standing hypothesis that the development of agriculture was an important selective pressure during recent human evolution.

Expanded summary*: The shift of human societies from a hunter-gatherer to an agrarian life style in the fertile crescent of Africa is presumed to have been the catalyst for the emergence of many major infectious diseases including tuberculosis, pertussis, small pox, influenza A, and measles to name a few. Yet, the extent to which hunter-gatherer and agricultural populations diverge in their immune response has not been evaluated, nor have we gauged the role of selection in contributing to these differences. We also know via recent selection studies that changes in regulatory regions, e.g. expression quantitative trait loci (eQTL), have played a dominant role in recent human evolution even more so than amino-acid altering changes. Yet, the contribution of eQTL in driving differences in innate immune response has not been thoroughly investigated. Elucidating these patterns can help us understand which regulatory changes have been pivotal in local adaptation to pathogens, and in doing so we can begin to hypothesize which classes of pathogens have been the most pertinent selection pressures, both of which are useful in vaccine development. In this study, we collected peripheral blood mononuclear cells (PBMCs) from both hunter-gatherer (Batwa) and agricultural (Bakiga) populations in Uganda offering a unique data set to evaluate this research question. We stimulated the PBMCs using viral (Gardiquimod–GARD) and bacterial (lippopolysacharide–LPS) ligands to mimicking infection and looked for a divergence in transcriptional regulation of innate immune response. We evaluated transcriptional differences between the Batwa and Bakiga populations (PopDE), and found 1,664 PopDE genes that differed in their overall expression following stimulation with LPS and 2,242 PopDE genes that differed in their overall expression following stimulation with GARD. Among these PopDE genes we found an increase in the anti-viral activity in the Batwa population with the increased expression of genes in interferon (IFN) pathways. We next mapped expression quantitative trait loci (eQTL) for 1,097 genes. We show that genes with cis-eQTL are enriched among PopDE genes, suggesting that a significant

fraction of transcriptional differences are genetically controlled. The higher expression of genes in the IFN-pathways found in the Batwa cannot be explained by cis genetic regulatory variants suggesting that either a trans eQTL or non-genetic factors are responsible for the increased activity of anti-viral responses. Finally, we show that several of the eQTL driving population differences in immune regulation have been targeted by recent positive selection reinforcing the long-standing hypothesis that the development of agriculture was an important selective pressure during recent human evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-350

Comparing the Statistical Fate of Paralogous and Orthologous Sequences

Florian Massip [1], Michael Sheinman [2], Sophie Schbath [3], Peter Arndt [4,*]

[1]Université Claude Bernard, Lyon, France, [2]Utrecht University, Utrecht, Netherlands, [3]Unit MIG, INRA, Jouy-en-Josas, France, [4]Max Planck Institute for Molecular Genetics, Berlin, Germany

Abstract: For several decades, sequence alignment has been a widely used tool in bioinformatics. For instance, finding homologous sequences with a known function in large databases is used to get insight into the function of nonannotated genomic regions. Very efficient tools like BLAST have been developed to identify and rank possible homologous sequences. To estimate the significance of the homology, the ranking of alignment scores takes a background model for random sequences into account. Using this model we can estimate the probability to find two exactly matching subsequences by chance in two unrelated sequences. For two homologous sequences, the corresponding probability is much higher, which allows us to identify them. Here we focus on the distribution of lengths of exact sequence matches between protein-coding regions of pairs of evolutionarily distant genomes. We show that this distribution exhibits a power-law tail with an exponent alpha = -5. Developing a simple model of sequence evolution by substitutions and segmental duplications, we show analytically and computationally that paralogous and orthologous gene pairs contribute differently to this distribution. Our model explains the differences observed in the comparison of coding and noncoding parts of genomes, thus providing a better understanding of statistical properties of genomic sequences and their evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-349

Transient Receptor Potential Gene Family Evolution in Invertebrates

Jun Gojobori [1,*]

[1]School of Advanced Sciences, SOKENDAI (The Graduate University for Advanced Studies), Hayama, Japan

Abstract: Transient receptor potential (TRP) superfamily are cation selective channels with six transmembrane domains. Many kinds of stimuli, such as chemical, mechanical, osmotic stress and heat can activate TRP channels. Some of the TRPs function as sensors for the environment and contributed to the adaptive evolution of the organisms. I conducted analysis to elucidate the whole repertories of TRPs in the genomes and transcriptomes of echinoderms and cnidarians. Based on HMM search using known TRPs, I found that echinoderms and cnidarians potentially have same subfamily members of TRP genes as vertebrates. I also found that sea urchin and coral have more TRPA genes than fruit fly or human. The unusual expansion of TRPA genes in them may have functionally important for their adaptation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-348

Adaptive process of visual systems from land to the sea in the sea snake

Takashi Seiko*

Abstract: Sea snake is one of a group in amniotae (mammals, reptiles and birds) have adapted completely to the aquatic environment from their terrestrial ancestors, although the terrestrial and sea environments are largely different from each other. Short and long wavelength of light is narrowed by scattering and absorbing as well as reduced light intensity. It was reported that the protein components (opsin) of visual pigments in a marine mammals, a group of amniotae which adapted perfectly to the sea, have adapted to such light environment. How did they adapt to the sea environment from land? I aim to reveal the adaptive process of visual systems associated with transition from land to sea in elapid snakes.
I extracted RNA from eyes of three types of snakes from Elapidae family (terrestrial, semi- and full-aquatic) and synthesize complementary DNA. Additionally I used genomic DNA from semi- and full-aquatic species. Using PCR, I amplified and identified nucleotide sequences of three opsin genes (*LWS*, *RH1* and *SWS1*) that was reported in a other snakes. and obtained the translated amino acid sequences. Then I compared these amino acid sequences between terrestrial and sea snakes. The opsins in sea snakes possess amino acid replacements that may tune the absorption of pigments. My results suggest that the aquatic adaptation from land is different between semi- aquatic and full-aquatic sea snakes. In addition, the result supports a possibility that amino acid changes found in sea snakes could play important roles in adaptation to the marine environment.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OT-OS7

The Genetic Identity of the Bangande People “The secret ones”

Hiba Babiker [1,*], Russell Gray [1]

[1]Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Jena, Germany

Abstract: Understanding human evolutionary history and past demographic events are becoming ripe for interdisciplinary research to complement genetics, linguistics, anthropological and historical studies. Their importance stems from previous research findings that Africa is the birthplace of modern humans. However, details of human prehistory in Africa remain largely obscure owing to the complex histories of hundreds of distinct populations.Therefore focusing on modern populations from Africa where ancient DNA is not accessible is a key to answering big questions in the history of our species. This project explores the genetic identity of the Bangande people who speak the Bangime language (Bangime is a language isolate spoken in the extreme Northwest of the Bandiagara Escarpment in Central Eastern Mali). We also aim at finding matches/mismatches between genetics and linguistics. We investigate 250 individuals from 12 West African populations representing different ethnicities and linguistic affiliations. Samples are genotyped with the Axiom® Genome-Wide Human Origins Array. Moreover and for the purpose of comparison, the final dataset combines publicly available datasets of African and non-African populations. Our analysis applies the most advanced statistical methods in population genetics. The outcomes of this project are critical for reconstructing African demographic history and also has insights on the coevolution of languages and genes. It highlights the importance of interdisciplinary research in decoding unanswered questions in the human history.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-351

Alignment averaging increases the accuracy of phylogeny inference

Haim Ashkenazy [1,*], Itamar Sela [2], Eli Levy Karin [1 3], Giddy Landan [4], Tal Pupko [1]

[1]Department of Cell Research and Immunology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel, [2]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, United States, [3]Department of Molecular Biology & Ecology of Plants, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel, [4]Institute of Microbiology, Christian-Albrechts-University of Kiel, Kiel, Germany

Abstract: The classic methodology of inferring a phylogenetic tree from sequence data is composed of two steps. First, a multiple sequence alignment (MSA) is computed. Then, a tree is reconstructed assuming the MSA is correct. Yet, inferred MSAs were shown to be inaccurate and alignment errors reduce tree inference accuracy. It was previously proposed that filtering unreliable alignment regions can increase the accuracy of tree inference. However, it was also demonstrated that the benefit of this filtering is often obscured by the resulting loss of phylogenetic signal. In this work we explore an *ad-hoc* approach, in which instead of relying on a single MSA, we generate a large set of alternative MSAs and concatenate them into a single super-MSA. By doing so, we account for phylogenetic signals contained in columns that are not present in the single MSA computed by alignment algorithms. Using simulations, we demonstrate that this approach results on average in more accurate trees compared to (1) using an unfiltered MSA; (2) using a single MSA with weights assigned to columns according to their reliability. Next, we explore in which regions of the MSA space our approach is expected to be beneficial. Finally, we provide a simple criterion for deciding whether or not the extra effort of computing a super-MSA and inferring a tree from it is beneficial. Based on these assessments, we expect our methodology to be useful for many cases in which diverged sequences are analyzed.

Expanded summary*: Dear committee,

Multiple sequence alignment (MSA) is often the first step in phylogeny inference. Furthermore, most current analyses rely on a single MSA, treating it as observed data. However, alignment methodologies produce MSAs which are subject to a considerable amount of errors. Thus, as previously shown, relying on any single MSA for downstream analyses in a bioinformatics pipeline is not advisable. Specifically, alignment errors were shown to reduce the accuracy of tree inference procedures.

The benefit of applying strategies to reduce the impact of alignment errors on subsequent analyses is highly debated. For example, previous studies suggested that filtering out unreliably aligned regions from the MSA prior to phylogeny inference can improve tree reconstruction accuracy. In contrast, it was demonstrated that the benefit of filtering is often obscured by the resulting loss of phylogenetic signal.

Recently, we have developed an alternative approach to account for alignment uncertainty when reconstructing phylogenetic tree. Our approach averages over uncertainty in MSA rather than to filter uncertain MSA regions. Specifically, alternative MSAs are generated by GUIDANCE2, thus accounting for uncertainty in MSA resulting from the assumed tree topology, the indel process, and co-optimal solutions. Comparing this methodology to standard filtering or weighting approaches, we show that we obtain significantly more accurate trees. Furthermore, we also characterize the cases in which averaging over MSAs is most beneficial. We expect our methodology to be useful and efficient for many cases in which relatively diverged sequences are analyzed.

Considering the high volume of citations for tree inference methods, our work is expected to be of special interest for the SMBE conference participants. I'd be very grateful if the committee supported my research by granting me a travel scholarship to present my work in the conference.

Sincerely,

Haim Ashkenazy

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS9

Genomic Insights into the Ancestry and Human Demography of Remote Polynesia

Alexander Ioannidis [1], Javier Blanco [2,*], Consuelo Quinto [2], Karla Sandoval [2], Erika Hagelberg [3], Mauricio Moraga [4], Thomas Parks [5], María Ávila-Arcos [6], Alexandra Adams [7], Celeste Eng [8], Esteban Burchard [8], Alexander Mentzer [5], Carlos Bustamante [7], Andrés Moreno-Estrada [2]

[1]Institute for Computational and Mathematical Enginnering, Stanford University, Palo Alto, United States, [2]Human Population Genomics Lab, Laboratorio Nacional de Genómica para la Biodiversidad, Irapuato, Mexico, [3]Department of Biology, University of Oxford, Blindern, Norway, [4]Facultad de Medicina, Universidad de Chile, Santiago de Chile, Chile, [5]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, [6]International Laboratory for Human Genome Research, Universidad Nacional Autónoma de México, Querétaro, Mexico, [7]Department of Genetics, Stanford University, Palo Alto, [8]Department of Medicine, University of California San Francisco, San Francisco, United States

Abstract: Beginning some three thousand years ago, the settling of Polynesia represents the final chapter in the expansion of humans across the globe. Although settled relatively late in historical terms, with occupation of the most remote islands occurring as recently as one thousand years ago, many questions remain about the peopling of this vast oceanic region. These questions include the sequence of island settlement, the dates of settlement, and the role of more recent admixture events in creating the modern island populations. Using dense genome-wide array genotyping of 445 modern samples from across the Polynesian archipelago, we attempt to answer some of these outstanding questions. In particular we investigate patterns of local ancestry within individuals, as well as patterns of relatedness within and across islands, to help elucidate historical settlement patterns. The widely separated Polynesian islands provide a uniquely structured canvas on which to implement novel variants of ancestry deconvolution techniques. We will describe the application of those techniques to human populations ranging from Near Oceania to Easter Island. Our results demonstrate the important role that both recent and ancient admixture events have played in creating the diversity pattern of modern Polynesian island populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-355

Identifying the genetic adaptations of salivary glands within Phyllostomidae

Michael Vandewege [1,*], Caleb Phillips [1]

[1]Texas Tech University, Lubbock, United States

Abstract: Any trait exposed to selection can be produced by a wide range of genetic causes and understanding the genetics of adaptation during species radiation is a significant field in evolutionary biology. Since the implementation of next-generation sequencing, it has become easier to sequence mRNAs from tissues to sample the genome in an impactful manner. Interestingly, there is evidence that salivary glands have played a major role in mammalian radiation because secreted products are recruited into several different biological processes. Further, salivary morphology and cellular structure are variable even among closely related species. To study the genes underlying the function of salivary glands, we sequenced the transcriptomes of submandibular glands within Phyllostomidae. In general, members within a mammalian family have generally similar diets. This is also true in bats where almost all are insectivorous, except for the megabats (Pteropodidae) which are obligate frugivores. In contrast the family Phyllostomidae is unique among mammals and bats because fundamentally discrete dietary strategies (insectivory, sanguivory, frugivory) are practiced within this family. We found that 5% of tested genes were evolving in an adaptive manner. Many of these proteins were involved in transmembrane regulation, cellular storage, cellular signaling and secretory pathways. Further, we found that proteins evolving adaptively among lineages were linked to dietary specializations. Here we present evidence that evolution at the molecular level mirrors the morphological and ultrastructural variation observed in this diverse family of mammals.

Expanded summary*: Identifying the genetic changes that enable ecological niche invasion is vital to understand the origins of biodiversity. However, it is a challenge in evolutionary biology to link genetic variation to natural selection and adaptation. Since the implementation of next-generation sequencing, it has become easier to sequence genomes and identify all genes evolving adaptively. Whole genome sequencing is still expensive, but sequencing the expressed mRNAs from tissues offers a cheaper alternative to sample the genome in an impactful manner. Therefore to study genetic adaptation, taxa that possess morphologically variable phenotypes or organs are helpful to identify adapting genes.

Salivary glands use ancient intracellular processes that involve the synthesis, modification and packaging of proteins in membrane-bound granules Mammals have three pairs of salivary glands: the submandibular, parotid, and sublingual glands. The submandibular gland shows wide anatomical variability among mammalian species which is often associated with dietary and other evolutionary specializations. Because submandibular glands show such wide morphological and structural variation even among close relatives, they are hypothesized to be directly involved in mammalian radiations and niche invasion.

To study the genes involved in the evolution and adaptation of salivary glands we have used a natural system opposed to a model. In general, members within a mammalian family occupy relatively similar niches and have similar diets. This is also true in bats where almost all are insectivorous, except for the megabats which are obligate frugivores. In contrast the family Phyllostomidae is unique among mammals and bats because almost all dietary strategies (i.e. insectivory, sanguivory, nectarivory, frugivory and carnivory) are practiced within this family. Phyllostomidae includes more than 200 species, began radiating approximately 30-35 MYA and it is proposed that most of the ecological divergence occurred within the first 10 MY of radiation. Because of these properties, phyllostomids offer an excellent system to study the evolution of genes expressed in submandibular glands. Given that phyllostomids radiated and adapted to novel niches relatively quickly, proteins involved in the adaptation process would also show evidence of

accelerated evolution. We sequenced submandibular gland transcriptomes of from nine phyllostomid species with discrete dietary strategies and two insectivorous outgroups.
We used d_N/d_S or ω ratio tests in PAML to identify proteins evolving under positive selection. We tested thousands of shared proteins expressed among all 11 species and found several trends. 1) Proteins used in transmembrane regulation, cellular storage, cellular signaling, and secretory pathways were often evolving adaptively. 2) We found that proteins adapting independently among lineages are linked to dietary specializations. Submandibular glands and their cellular ultrastructure are morphologically variable and we have found that cellular structure and storage proteins are rapidly evolving, thus identifying a genetic basis for this variation, thus linking genetic and morphological adaptation. Salivary glands are overlooked in biology, however their role in evolution may be larger than recognized.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-353

Evolutionary analysis of the IFI16 like (IFI16l) PYHIN family gene in Indian Ruminants

Sushil Kumar on behalf of Animal Genomics Lab, Animal Biotechnology Centre, National Dairy Research Institute Karnal, India, Ashutosh Vats, Jatinder Chera*, Sachinandan De and Animal Genomics Lab, Animal Biotechnology Centre, National Dairy Research Institute Karnal, India

Abstract: AIM2 like Receptors (ALRs) are germline encoded Pattern Recognition Receptors (PRRs) that act as a cytoplasmic DNA sensor molecules in case of any viral or bacterial infection resulting in induction of type I interferon (IFN) and other proinflammatory cytokines. They belong to PYHIN gene family of cytokines, characterized by having N-terminal PYRIN (PYD, PAAD or DAPIN) domain and C-terminal HIN-200 (hematopoietic, interferon inducible nuclear protein with 200 amino acid repeat) domain joined by a linker region. They also contain an N-terminal nuclear localization signal.

The syntenic organization of the PYHIN gene cluster region was studied in important mammalian species. This genomic region is flanked by cell adhesion molecule 3 (CADM3) gene and spectrin alpha chain (SPTA1) gene. This genomic region is very dynamic with variable lineage specific expansion. In ruminants only one functional ALR genes was found compared to five in human, thirteen in mice. Surprisingly, among mammals, bats of order chiroptera are not having any ALR gene.

To understand the intracellular DNA sensing in Indian domestic ruminants, we characterize the IFI16l gene in Indian cattle (*Bos indicus*), water buffalo (*Bubalus bubalis*), sheep (*Ovis aries*) and goat (*Capra hircus*). Ruminant IFI16-like gene contains 521 amino acid residues in Indian cattle, water buffalo and sheep while 519 residues in the case of goat. There are three isoforms of IFI16l gene (A, B and C) in each of the studied animals with different amino acid length due to variable length of corresponding inter-domain region. In Indian cattle IFI16l-C isoform lacks PYD domain while C-isoforms of water buffalo, sheep and B-isoform of goat lacks the HIN domain. The HIN domain binds with altered self and pathogenic DNA molecules while PYD domain involved in homotypic interaction for downstream signaling. The alternative splicing producing such types of POPs (Pyrin only proteins) and HOPs (HIN only proteins), transcript variants of IFI16l gene might be playing significant role in intracellular immune response in absence of other ALR family members.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS3

Genetic manipulation of entire populations with CRISPR gene drives

Jackson Champer [1], Robert Unckless [2], Andrew Clark [1], Philipp Messer [1,*]

[1]Cornell University, Ithaca, [2]University of Kansas, Lawrence, United States

Abstract: A functioning “gene drive” system could fundamentally change our strategies for the control of vector-borne diseases such as malaria, dengue, yellow fever, and Zika, by allowing us to drive genetically engineered alleles into vector populations. Such alleles could prevent pathogen transmission by the vectors or alter life history traits to reduce vector capacity. The recently developed CRISPR/Cas9 gene drive system promises a highly adaptable mechanism for this purpose that works by converting heterozygotes for the driver construct into homozygotes in the germline. However, it remains unclear how well this mechanism would work in natural populations where the evolution of resistance could thwart the spread of a driver. We present both theoretical and experimental results that shed light on how such resistance alleles can emerge during the drive process, how genetic variability in the population impacts their formation rate, and how these alleles are expected to ultimately affect the spread of a driver in the population. We show that the key factor determining the probability that resistance evolves is the formation rate of resistance alleles, while the conversion efficiency of the driver, its fitness cost, and its introduction frequency have only minor impact. Our experiments in the model system *Drosophila melanogaster* confirm that such resistance alleles arise frequently during incomplete driver conversion in the germline, as well as after fertilization in the embryo due to persistence of maternal Cas9 or “leaky” Cas9 expression. These results inform strategies that could facilitate the engineering of drivers with lower resistance potential, and motivate the possibility to embrace resistance as a possible mechanism for controlling a drive.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OT-OS8

Human X chromosomes show recurring selective sweeps suggestive of meiotic drive

Kasper Munch [1,*], Elise Lucotte [1], Moises Coll Marcia [1], Mikkel H Schierup [1]

[1]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark

Abstract: The unique inheritance pattern of the primate sex chromosomes makes them susceptible to very strong selection in the form of meiotic drive. Strong selective sweeps have differentially shaped X chromosome diversity in great ape species and across the evolutionary time scales represented by their common ancestors. The locations of regions where diversity has been depleted by sweeps vary dramatically between populations and species, but also reveal regions repeatedly targeted by sweeps across evolutionary time scales. These regions are enriched for ampliconic genes post-meiotically expressed in testis suggesting that the force driving these sweeps are meiotic drive resulting from inter-chromosomal conflict between the X and Y chromosomes for transmission to the next generation. Here we show that diversity in these regions of the X chromosome also vary dramatically across human populations. We find wide regions depleted of variation each shared by subsets of human populations and use ARGweaver deliniate and date putative sweeps. Our results suggest that sweeps responsible for these patterns have occurred during the human colonisation of the world.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OM-OS3

Parallel selective sweeps of selfish Segregation Distorter complexes in African and European Drosophila melanogaster populations.

Amanda Larracuente*, Cara Brand [1], Daven Presgraves [1]

[1]Biology, University of Rochester, Rochester, United States

Abstract: Segregation distorters are selfish genetic elements that unfairly achieve biased transmission through the germline. Found across a wide variety of taxa including fungi, plants, insects and mammals, segregation distorters can rapidly increase in frequency in natural populations and trigger the evolution of suppressors across the genome. One of the best-studied male segregation distorters is the autosomal *Segregation Distorter* (*SD*) gene complex of *Drosophila melanogaster*. Males heterozygous for *SD* and a wild type chromosome transmit *SD* to over >95% of their progeny, whereas heterozygous females transmit *SD* fairly to 50% of their progeny. *SD* is found at frequencies of ~1-5% in natural populations across the globe. Cosmopolitan *SD* chromosomes involve a main driving locus on chromosome *2L—Sd-RanGAP—*and several upward enhancers that strengthen drive on both *2L* and *2R*. *SD* chromosomes often tighten linkage among enhancers and *Sd-RanGAP* via the recruitment of chromosomal inversions that suppress recombination. Suppressed recombination prevents the distorter from recombining onto a sensitive target background and generating self-distorting "suicide" genotypes. African and European *SD* chromosomes appear molecularly different—they do not share inversions and we find that suppressors of European *SD* chromosomes do not suppress African *SD* chromosomes. While inversions provide short-term benefits to *SD*, the reduced recombination entails long-term costs due to the associated reduced efficacy of natural selection. To study the consequences of suppressed recombination, we performed whole-chromosome population genomics analyses of *SD* chromosomes sampled from African and European populations. We Illumina-sequenced 10 haploid embryos from Zambia and 10 adults from France. We use these data to study the patterns of nucleotide diversity and the accumulation of potentially deleterious mutations on *SD* chromosomes. We find a dearth of nucleotide variation on French *SD* chromosomes that begins at *Sd-RanGAP* on chromosome *2L*, spans the centromere and extends 4 Mb into chromosome *2R*. In contrast, we find a more dramatic dearth of nucleotide variation on Zambian *SD* chromosomes that begins at *Sd-RanGAP* on chromosome *2L*, spans the centromere and extends for ~22.5 Mb across chromosome *2R*— a massive sweep signal that suggests very recent and strong selection. We therefore detect a signature of parallel sweeps of independent multi-locus selfish *SD* complexes in Africa and France. Taken together, differences in (i) the structure of the selective sweeps, (ii) the chromosomal inversions involved, and (iii) responses to genetic suppressors of *SD*, imply that Zambian and French *SD* gene complexes have functionally diverged from one another.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-335

MUTATION RATES ASSOCIATED WITH DNA TRANSPOSONS IN MYOTIS

Nicole Paulat [1,*], David Ray [1]

[1]Biological Sciences, Texas Tech University, Lubbock, United States

Abstract: Transposable elements (TEs) are DNA sequences that mobilize through copy-and-paste or cut-and-paste mechanisms, expanding within a host genome. *Myotis* is one genus within vespertilionid bats which has experienced an unorthodox TE history. For example, their genomes are unique among mammals in containing many active DNA transposons, which continue to shape their genomic landscapes. Recent data suggests that, in addition to the indel mutations normally associated with TE activity, these genetic elements may also contribute to higher mutation rates via DNA repair mechanisms. DNA transposons preferentially insert near genes, and so transposon activity may be correlated with mutation rate increases in regulatory regions and coding sequences. An analysis of transposon polymorphisms in nine *Myotis* species will reveal the extent of mutations in nearby genes that are associated with DNA repair after transposon insertions and excisions. These increased mutation rates could correlate to differences in orthologous genes between closely related *Myotis* species and contribute to our understanding of this exceptionally diverse clade.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-366

Optimized Circular Sequencing Approach Reveals Universally High Transcription-Error Rates in Bacteria

Weiyi Li [1,*], Michael Lynch [1]

[1]Biology, Indiana University Bloomington, Bloomington, United States

Abstract: Errors can occur at any level during the replication and expression of genetic information. Genetic mutations, which are derived mainly from replication errors, have been extensively studied in evolutionary research. However, many fundamental details, such as the rate, spectrum and selective constraints of transcript errors remain largely unknown. Current information on transcript errors largely relies on reporter-construct assays, which only focus on individual loci and cannot identify errors without phenotypic effects. Recently, two high-throughput methods, replicated sequencing (Rep-seq) and circular sequencing (CirSeq), have been proposed to identify genome-wide transcriptional errors *in vivo*. However, both of these methods can cause a significant amount of RNA damage during library preparation. Here, we optimized the CirSeq approach and applied it to *Escherichia coli*, *Bacillus subtilis*, *Agrobacterium tumefaciens* and *Mesoplasma florum*. Analysis of more than 7,000 transcriptional errors indicates universally high transcription-error rates in bacteria, which are 4 to 5 orders of magnitude higher than the corresponding genetic mutation rates. Intriguingly, different from other three bacterial species, *Mesoplasma florum* shows a spectrum of errors dominated by G-to-A substitutions instead of C-to-U, which can result from the deamination of cytosine. We also found that the error rate of non-coding RNAs is significantly lower than that of mRNAs, suggesting a stronger selective constraint on functional transcripts that are not translated into proteins.

Expanded summary*: Transcript errors can directly cause dysfunctions considering the regulatory roles of small RNAs and the fate determination of mRNAs by RNA structure motifs. These errors can also indirectly give rise to misfolded proteins and induce proteotoxic stress. Thus, transcript errors can impose a load on cellular integrity and represent a new molecular mechanism to influence fitness.

Here, we optimized the recently proposed circular sequencing (CirSeq) approach and provided an unbiased evaluation of genome-wide transcript errors *in vivo*. Our result reveals universally high transcriptional-error rates in bacteria and sheds light on the rate, spectrum and selective constraints of transcript errors.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-341

Establishing The Genetic Basis Of Hibernation In The 13-Lined Ground Squirrel

Katharine Grabek [1,*], Carlos Bustamante [1]

[1]Genetics, Stanford School of Medicine, Stanford, United States

Abstract: Hibernation, extraordinarily dynamic and extreme for mammals, challenges current understandings of homeostasis at cell, tissue and whole-body levels. During the winter, mammalian hibernators enter into a state of torpor, whereby physiological processes are dramatically reduced, and body temperature is lowered to near freezing. However, torpor is not continuous during this 6-9 month period of hibernation, but is instead punctuated by brief, but metabolically intense, arousals back to basal physiology. Although the basic ecology, anatomy and physiology of hibernators have been well-studied, the underlying genetic components that drive these cycles of torpor and arousal, along with the annual cycle of hibernation, remain poorly understood. Traditional genetic studies linking genome to phenome are notably lacking in hibernation, with almost nothing known about genetic variation within hibernators. The goal of this project is to establish the genetic basis of hibernation in our model hibernator, the 13-lined ground squirrel, *Ictidomys tridecemlineatus*. To this end, we first develop a high-quality genomics resource for this species. This includes bringing the current genome assembly to chromosome scale by adding long-range sequence data from Dovetail Genomics. Here, we nearly triple the scaffold N50 from 8Mb to approximately 23Mb. Using a genotype-by-sequencing strategy, we also construct a dense linkage map and characterize genetic variation in individuals from full-sibling families. We recover several hundred thousand variants located genome-wide, allowing us to detect significant population substructure. Finally, our newly created resource enables us to estimate the heritability of, and identify variants associated with, hibernation-related traits that have been measured from body temperature telemeters.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-375

Pedigree Inference Using Markov Chain Monte Carlo

Amy Ko*, Rasmus Nielsen

Abstract: The pedigree of a sample of individuals can provide valuable information in a wide range of genetic studies including linkage analyses, heritability estimation, detection of selection, and demographic inferences. In particular, using the fine-resolution genealogical relationships among individuals provided by the pedigree has a potential to improve demographic inference of the very recent past, for which coalescent-based methods may not work well. Despite the importance of pedigree inference, existing methods are limited to inferring only close relationships or analyzing a small number of individuals or loci. We present a Markov chain Monte Carlo (MCMC) method for estimating pedigrees in large samples of otherwise seemingly unrelated individuals. The method supports complex pedigree structures such as polygamous families, multi-generational families (up to 5 generations), and pedigrees in which many of the member individuals are missing. Computational speed is greatly enhanced by the use of a composite likelihood function which approximates the full likelihood. Using simulations, we show that the new method leads to improved estimates of relatedness compared to both existing pedigree inference methods and pairwise relatedness estimation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-346

Insights into genome evolution in the ciliate class Karyorelictea (i.e. relict nucleus) through single-cell 'omics' and quantitative PCR

Ying Yan*, Xyrus Maurer-Alcalá [1], Laura Katz [1]

[1]Smith College, Northampton, United States

Abstract: Karyorelictea are an unusual class of ciliates that have been argued to have "paradiploid" somatic macronuclei (i.e. each protein coding gene has ~2 copies; Raikov and Kovaleva, 1978). Karyorelictea are unique among ciliates in that they possess non-dividing somatic macronuclei and instead must generate new macronuclei from germline micronuclei within each cell division. Despite these unusual features, very limited data are available for Karyorelictea (e.g. only eight protein sequences are available on GenBank). By combining the light microscopy, quantitative PCR and single cell 'omic' techniques, we are exploring the genome structure of this group of uncultivable ciliates. Our preliminary data suggest that differential amplification of protein coding genes occur in Karyorelictea. Comparisons of HTS data generated from both genomes and transcriptomes of single cells yields a wealth of data on gene family evolution in this class, and provides preliminary insights into genome structure. Together, our data indicate that somatic nuclei in Karyorelictea are neither primitive nor "paradiploid". Instead, Karyorelictea represent an independent experiment in the evolution of distinct germline and somatic genomes.

Expanded summary*: As a group of microbial eukaryotes, ciliates are characterized by dimorphic nuclei (i.e. germline nuclei and somatic nuclei) and the presence of cilia in at least one of life stages. Numerous key discoveries have been made through studies of ciliates, including the discovery of self-splicing RNA and telomeres/telomerases. Yet the bulk of the work on genome structure and nuclear organization of ciliates have been focused on model ciliates such as *Tetrahymena* and *Paramecium*.
The ciliates in the class Karyorelictea are unique by their non-dividing somatic nuclei that are argued to be "paradiploid". Karyorelictea differentiate new macronuclei from micronuclei with each cell division and an individual macronucleus only persists for a few generations before degrading. Due to our current inability to maintain karyorelictids in culture, only few studies have been characterized on nuclear features of this group of ciliates. By taking advantage of single cell "omic" technique and quantitative PCR, we are able to revalue the "paradiploid" hypothesis, which is suggested by Raikov and Kovaleva (1978) based on morphological and cytochemical approaches, and gain more insights into genome structure of Karyorelictea ciliates. Karyorelictea ciliates have always been considered primitive, which is not consistent with our preliminary data. Furthermore, given the non-dividing somatic nuclei, karyorelictids are one of the key groups to unveil the genomic evolution of ciliates. Our preliminary HTS data will largely enrich the diversity of ciliate database, thus will bring us new insights in many aspects of ciliates genomic evolution such as gene family evolution and copy number regulation. Thus, our work represents not only a case study for using modern techniques to reinvestigate previous hypothesis, but also a study of a unique member in eukaryotes with distinct germiline and somatic genomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-331

Kin selection drives the molecular evolution of pre-stalk cells in the social amoeba

Katherine Geist [1,*], Suegene Noh [1], Joan Strassmann [1], David Queller [1]

[1]Biology, Washington University in St. Louis, Saint Louis, United States

Abstract: The social stage of the amoeba, *Dictyostelium discoideum*, is triggered when solitary cells begin to starve. Cells then develop into a multicellular fruiting body in which reproductive spores are held aloft a sterile stalk. Stalk cells die to create a structural support that likely aids the dispersal of spores. This suggests that by helping their relatives to disperse to more favorable locations, altruistic sacrifice of stalk cells is favored by kin selection. However, for microbial cooperative traits, like the differentiation of cells into stalk at the expense of their own reproduction, it can be unclear how important that trait is in the wild. Are stalk cells gaining benefit indirectly through highly-related spores? Or as has been challenged, are stalk cells simply "making the best of a bad job" and under selection to reproduce directly? We test the two competing hypotheses by comparing molecular evolution signatures of kin selection vs. conditional selection. Specifically, we compare the strength of selection on deleterious mutations in pre-stalk vs. pre-spore genes. In support of the kin selection hypothesis, we find the strength of purifying selection to be roughly equal between pre-stalk and pre-spore cells.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-388

Characterization and molecular evolutionary analysis of a newly discovered selfish X chromosome in Drosophila

Graeme Keais [1,*], Mark Hanson [1], Brent Gowen [1], Steve Perlman [1]

[1]Biology, University of Victoria, Victoria, Canada

Abstract: Selfish genetic elements are widespread and powerful forces in evolution. These genomic parasites increase their own transmission relative to the rest of the genome, thereby spreading rapidly in populations even if they contribute negatively to the fitness of their host. Driving X chromosomes (X^D) are a type of selfish genetic element found in a wide range of taxa, including mammals, plants, and insects. In insects, these selfish chromosomes bias their transmission by destroying or incapacitating Y-bearing sperm. As a result, males that carry an X^D transmit almost exclusively X-bearing gametes, and therefore produce predominantly female offspring. X chromosome drive often instigates intragenomic conflict, and has been shown to significantly affect chromosome organization, the evolution of mating systems, and patterns of molecular evolution. Furthermore, an unhampered increase in X^D frequency could drive a species to extinction due to a lack of males. We have characterized several features of a newly discovered X^D in a common and wide ranging mushroom-feeding Drosophila species, *Drosophila testacea*. For example, we demonstrate that males carrying the X^D sire between 80-100% female offspring, and most of their sons (of which there are few) are sterile, and appear to lack a Y chromosome. We will also present molecular genetic analyses demonstrating that the X^D in *D. testacea* is old – potentially older than the species itself – and that it exhibits interesting sequence variation as a consequence of reduced recombination.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-343

An innovative cancer classification method with tumor-educated blood platelets

Guangzao Huang [1,*], Moliang Chen [1], Fatemeh Karimidehcheshmeh [1], James Cai [2], Guoli Ji [1]

[1]Xiamen University, Xiamen, China, [2]Texas A&M University, College Station, United States

Abstract: Molecular profiling of tumor tissue samples has emerged as possible cancer classifying method. Tumor-educated blood platelets (TEPs) are implicated as central players in the systemic and local responses to tumor growth, thereby altering their RNA profile. It is reported that mRNA profiles of tumor-educated blood platelets enable for cancer classification. We present a new classification algorithm, which addresses the classification problem with reliability analysis. This method integrates multiple fitting regression and Bayes Decision theory, called MFR-B. Multiple fitting regression is initially put forwarded for spectra multivariate regression analysis. This method is demonstrated to be a useful tool for linear and nonlinear multivariate regression analysis. Here, we use multiple fitting regression for discriminant analysis, which uses the features as the independent variables and the class label as the dependent variable. Then the predicted values of samples given by multiple fitting regression are used as a feature extracted from the TEP RNA profiles. With the good fitting capacity of multiple fitting regression, the extracted feature fluctuates around the class label and the samples are clustering in the one-dimensional space, which makes the distribution of each class more easily to be described. With the estimated probability density function of each class and Bayesian decision theory, a Bayesian classifier is built, which is optimal with respect to minimizing the classification error probability. Bayesian classifier makes classification decision by maximizing posterior probability, which measures the reliability of classification result. The dataset of TEP RNA profiles and a simulated dataset are used to verify the effectiveness of the MFR-B model. Two state-of-the-art methods, SVM and PLS-DA are also applied to these datasets to achieve a comprehensive comparison. The results show that MFR-B is favored over SVM and PLS-DA for cancer diagnosis with TEP RNA profiles. In particular, MFR-B can provide more proper reliability measure for the diagnosis result than SVM and PLS-DA.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-327

Recombination in sticklebacks identified via whole-genome pedigree resequencing

Jason Sardell [1,*], Changde Cheng [1], Andrius Dagilis [1], Asano Ishikawa [2], Matthew Josephson [3], Jun Kitano [2], Catherine Peichel [3], Mark Kirkpatrick [1]

[1]Department of Integrative Biology, University of Texas at Austin, Austin, United States, [2]Department of Population Genetics, National Institute of Genetics, Shizuoka, Japan, [3]Institute of Ecology and Evolution, University of Bern, Bern, Switzerland

Abstract: Recombination plays an important role in generating genotypic diversity, but studies of crossover recombination have been limited to a handful of model organisms. We undertook the most comprehensive analysis to date of recombination in teleost fish, based on whole-genome sequencing of 15 parent-offspring quartets of two sister species of sticklebacks (*Gasterosteus aculeatus* and *G. nipponicus*). We used identity by descent to phase individual haplotypes. We then identified locations of recombination breakpoints for each transmission event. Recombination breakpoints were enriched for GC content and *Prdm9* binding motifs, indicating that this gene may mediate recombination in teleost fish. In contrast to other systems, recombination breakpoints were randomly distributed with regard to gene location and transcription start sites. Even so, we found evidence for recombination hotspots in the genome. Additionally, recombination rates were higher in females than males. Recombination breakpoints also were enriched towards the telomeric ends of chromosomes in males but not females. These results provide unique insights into the evolution of crossover recombination among vertebrates.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OT-OS4

New World Africans are an amalgamation of African peoples with modest gene flow from non-Africans

Fatimah Jackson [1,*] and Genomic Models Research Group

[1]Howard University, Washington, United States

Abstract: Genomics link African-descended peoples on all sides of the Atlantic and throughout the Americas. New World Africans are the descendants of survivors of the forced migraitons of Africans to the Americas during the 16th through 19th centuries. During this time period, up to 20 generations of descendants have been exposed to a wide range of novel as well as reoccurring selective pressures within the American context. Genomically, today this heterogenous population is an amalgamation of African peoples, primarily as a consequence of intra-African gene flow, mainly from West and West Central Africa, with limited Southeast African contributions and modest, regionally-defined gene flow with specific non-Africans. New World Africans show variable geospatial and temporal patterns of genomic diversity that are amplified by regional patterns of non-genomic variations (e.g., microethnic group affiliations). Using selected autosomal markers, we hypothesize that current genomic diversity ajmong New World Africans counters historic variance in African regional origins, reflects regional shifts in selective pressures associated with specific migrations, and unique demographic opportunities in the Americas for genetic drift. Using the models of ethnogenetic layering, we depict the population biology of African peoples in the Americas over historical time and the regional transformaions in various subsets.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS15

Signatures of constraint on mammalian pseudogenes functioning as competitive endogenous RNAs

Cian Glenfield [1,*], Aoife McLysaght [1]

[1]Smurfit Institute of Genetics, Trinity College, University of Dublin, Dublin, Ireland

Abstract: Competitive endogenous RNAs (ceRNAs) represent a novel class of post-transcriptional gene regulators, which function by competing with other RNA transcripts that have microRNA (miRNA) binding sites in common. Expressed pseudogenes and other long non-coding RNAs have been shown to alter the levels of mRNA from protein-coding genes by acting as ceRNAs, and improper regulation of ceRNA expression can affect cancer development and progression. BRAFP1, a processed pseudogene originating from the BRAF gene, functions as a ceRNA that plays a role in certain cancers in both human and mouse cells, but appears to have evolved independently in these lineages. Despite these findings, few studies to date have considered the evolutionary impact and origins of ceRNAs. Here we report our results on the evolutionary analysis of the BRAF pseudogene in human (BRAFP1) and mouse (Braf-rs1). We find that BRAFP1 is present in syntenic locations in each species of the Catarrhini lineage. Multiple sequence alignment and substitution analysis reveals that the 3'UTRs of the pseudogenes have a lower substitution rate relative to their pseudo-protein-coding regions, and a similar rate to their respective parent gene 3'UTRs. In addition, we found several miRNA binding sites that appear conserved between BRAF and BRAFP1 for miRNAs previously validated to regulate both transcripts in human cells. With respect to Braf-rs1 we find that this pseudogene is species specific, and differs minimally between strains. Our results provide a reconstruction of the evolutionary history of BRAFP1 and also explore the potential for functional conservation of these novel regulatory elements.

Expanded summary*: The competitive endogenous RNA (ceRNA) hypothesis has attracted much interest and controversy in recent years owing to its potential as a unifying theory on the function of seemingly non-functional non-coding RNA transcripts [1,2,3]. This theory proposes that RNA transcripts, including long non-coding RNAs, circular RNAs and protein-coding mRNAs, are capable of indirectly regulating one another via their shared microRNA (miRNA) recognition elements (MREs). Several studies have shown how up- or downregulation of certain ceRNAs, which comprise pseudogenes and other long non-coding RNAs, can influence the development and progression of cancer [1,4,5]. One such study showed that a processed pseudogene, BRAFP1, originating from the BRAF gene, is capable of regulating the transcript levels of its parent gene in both human and mouse cell lines [6]. Sequence analysis indicated that the human and mouse pseudogenes are not orthologous. Overexpression of the pseudogene resulted in the development of an aggressive malignancy similar to human diffuse large B cell lymphoma. Conversely, silencing of pseudogene expression was associated with a reduction in BRAF mRNA levels and in the rate of cellular proliferation. The convergent evolution of these non-orthologos BRAF pseudogenes in two distinct species to perform the same ceRNA function prompted us to investigate this relationship further and reconstruct their evolutionary history.

Here we show that by querying human BRAFP1 against other primate and mammalian genomes using BLASTN we discovered the presence of further BRAFP1 pseudogenes in 8 additional species of the Catarrhini lineage. The absence of BRAFP1 in other mammals outside of this lineage suggests a single origin of this pseudogene, dating to just after the Old world monkeys and Apes diverged from New world monkeys. Applying the same method for the mouse Braf-rs1 pseudogene, we found no orthologs in any other rodent or mammal lineages tested, indicating mouse lineage-specificity. We confirmed these results through multiple sequence and phylogenetic tree analysis of the Catarrhini-specific BRAFP1, mouse Braf-rs1 and their BRAF parent genes. Interestingly, we also found that the 3'UTRs of BRAFP1 have a similar sustitution rate compared to the 3'UTRs of their respective BRAF genes, but a lower substitution rate than their pseudo-protein-coding regions, indicating potential sequence constraint on this region. Additionally, our MRE analysis revealed several binding sites that are present in both BRAF and BRAFP1 in each species tested. Using publically available RNA-Seq datasets we have detected low expression levels of BRAFP1 in some normal tissue types in human and additional primates and low levels of Braf-rs1 in mouse, suggesting BRAF pseudogene expression may be cell type specific. These results suggest that a functional BRAF pseudogene has arisen independently at least twice in mammals. The relevance of ceRNAs to these events is suggestive, and our studies aim to further scrutinize this hypothesis.

Since disruption of ceRNA activity can lead to aberrant cellular behaviour and human pathologies such as cancer [1,4,5,6], we expect these ceRNAs to exhibit some amount of dosage constraint. The potentially conserved role of ceRNAs may have important implications for dosage balance and sensitivity of these novel regulatory elements in the genome throughout evolution. Conversely, a

lack of constraint acting on ceRNAs could demonstrate that these are lineage-specific adaptations, such as Braf-rs1 in mouse. Given the current lack of research on the evolution of ceRNAs we expect that our insights will help facilitate further analysis and understanding of these regulatory elements.

References

[1] Poliseno Let al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465(7301): 1033-1038 (2010)

[2] Salmena L, Poliseno L, Tay Y, Kats L & Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language. *Cell* 146(3): 353-358 (2011)

[3] Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. *Nature Reviews Genetics* 17: 272-283 (2016)

[4] Zheng L et al. Competing endogenous RNA networks of CYP4Z1 and pseudogene CYP4Z2P confer tamoxifen resistence in brease cancer. *Mol Cell Endocrinol.* S0303-7207(16)30060-0 (2016)

[5] Yu F et al. MALAT1 functions as a competing endogenous RNA to mediate Rac1 expression by sequestering miR-101b in liver fibrosis. *Cell Cycle* 14(24): 3885-3896 (2015)

[6] Karreth FA et al. The BRAF pseudogene functions as a competetive endogenous RNA and induces lymphoma in vivo. *Cell* 161(2): 319-332 (2015).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS13

How brown rats adapted to life in NYC's concrete jungle

Arbel Harpak [1,*], Nandita Garud [2], Dmitri A. Petrov [1], Noah A. Rosenberg [1], Pleuni S. Pennings [3], Jason Munshi-South [4]

[1]Stanford University, Stanford , [2]University of California San Francisco, [3]San Francisco State University, San Francisco, [4]Fordham University, Armonk, United States

Abstract:

Brown rats (*Rattus Norvegicus*) have recently grown to enormous population sizes in urban environments. As a result, they are responsible for billions of dollars in damage to human health and infrastructure. What role did genetic adaptation play in the spread of rats in cities? To answer this question, we collected whole-genome samples from 29 brown rats from New York City (NYC) and scanned for genetic signals of adaptation. We applied multiple genomic tools, each tailored to identify specific modes of adaptation. Intriguingly, we found evidence for recent selective sweeps in groups of genes associated with sensory perception and regulation of proteolysis. We investigated whether these adaptations are specific to urban environments by scanning for similar adaptations in a sample of brown rats living in their presumed ancestral range in rural north-east China. Finally, to elucidate differences and similarities to artificial selection in domesticated animals, we compare our inferred adaptations in these wild urban rodents to adaptations previously associated with domestication, tameness and aggression.

Expanded summary*:

Brown rats are responsible for billions of dollars in damage to human health and infrastructure. One analysis estimated that 1.25 billion rats inhabit the United States, causing $19 billion of damage annually (Pimentel et al. 2000). However, we know relatively little about the ecology and evolution of rats (Feng & Himsworth 2014). This lack in knowledge is striking given that brown rats are wild conspecifics of the laboratory rat, which gives us the possibility to use genomic resources from the laboratory rat (Gibbs et al. 2004) to study wild brown rats. In addition, brown rats are ubiquitous in New York City (NYC), one of the most dense and well described urban areas in the world, which makes brown rats an ideal system for urban landscape genomics over a fine-grained spatial scale. Furthermore, rats are a powerful model for investigating historical demography and natural selection because of the available genomic resources, their global distribution, the putative intensity of selective pressures on their urban populations and the availability of large numbers of museum specimens. As a model system, rats are therefore rivaled by only a small list of vertebrates, few of which have such large wild-living populations and have not experienced intense artificial selection through domestication.

This model organism / population allows us a unique prism to key questions of natural selection. How big of a role did adaptation play in the spread of rats in NYC? If such adaptive events can be spotted in contemporary genomes, what is the genetic mode of these adaptations? Finally, what are these adaptations?

In this study, we develop brown rats (AKA Norway or sewer rats, *Rattus norvegicus*) as a model system for examining the influence of urbanization on the evolutionary dynamics of a wild-living mammal. We use whole-genome sequencing to detect signatures of recent natural selection in NYC rats. Specifically, we identify genes and pathways that are foci of recent selection and characterize the modes (e.g. hard vs. soft sweeps) and genomic features of recent adaptive events. In addition, we develop spatial analysis tools to examine the relationship of genetic adaptation in this wild population and the cityscape.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-407

Recent admixture and migration dynamics of post-Columbian Mexico

Juan Esteban Rodriguez-Rodriguez [1,*], Javier Blanco-Portillo [1], Alexander Ioannidis [2], Andres Moreno-Estrada [1]

[1]Human Population Genomics Lab, LANGEBIO, Irapuato, Mexico, [2]Institute for Computational and Mathematical Engineering, Stanford University, Palo Alto, United States

Abstract: Mexico has a considerable population substructure due to historical events anddistinct amounts of admixture between ethnic groups such as Africans, Europeans,Native Americans and to a lesser extent East Asians. Using genome-wide SNP array datafrom indigenous and admixed Mexican populations, we are exploring the ancestry tractlength distribution from 7 different states across Mexico to infer the timing ofadmixture in each region. To that end we applied Tracts, a method that examines thedistribution of all ancestry length blocks in a given admixed population. We arealso applying ancestry-specific approaches to better pinpoint the origin ofunderstudied components such as the sub-continental origin of Asian haplotypes indifferent regions across Mexico. We expect to relate these admixture and migrationdynamics with historical events probably related to conflicts, slavery, epidemics,natural disasters or other phenomena in the past few centuries.

Expanded summary*: As is the case with many Latin American countries, Mexico's history is not uniform across its states. The interaction of ethnic groups have varied greatly among regions, for instance, native groups like the Mayan people in Yucatan, as well as non-native groups like Africans in Veracruz and Guerrero. Many of these groups have left a genetic footprint in modern Mexican populations. My goal is to identify these differences between states in the north, center, south and southeast of the country, in order to pinpoint the subcontinental nature of each of the ancestries found within Mexico. Additionally, I am interested in estimating how the admixture process occurred in terms of timing and the number of migration waves.

This research is fascinating and relevant because there are still questions to explore about Mexican's ancestry, beyond the superficial view of it being merely an admixture of Native Americans, Europeans and Africans. Biomedical research demands the whole story to be taken into account, as historical records suggest there could be more ancestries like Italian, Filipino, Chinese, among many more. Furthermore, Native American genetic profiles are far from being homogeneous, and it has also been observed that Native American ancestry in admixed Mexicans have a geographical pattern. There could be a genetic predisposition in Mexicans from a specific state or city due to an unexpected ancestry contribution. Finally, this kind of analysis will help us elucidate the origin of our ancestors and unravel history.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-332

Hierarchical social networks shape gut microbial composition in wild Verreaux's sifaka

Amanda Perofsky [1,*], Rebecca Lewis [2], Laura Abondano [2], Anthony Di Fiore [2], Lauren Meyers [1 3]

[1]Integrative Biology, [2]Anthropology, The University of Texas at Austin, Austin, [3]Santa Fe Institute, Santa Fe, United States

Abstract: In wild primates, social behavior influences exposure to environmentally acquired and directly transmitted microorganisms. Prior studies have shown that gut microbiota reflect pairwise social interactions among their primate hosts. We demonstrate that higher order social network structure, beyond pairwise interactions, drives gut bacterial composition in a wild population of Verreaux's sifaka (*Propithecus verreauxi*). Membership in seven social groups clearly shaped microbiome diversity and similarity between individuals. Within social groups, dominant males and gregarious individuals harbored the most diverse microflora. Thus, social groupings likely shape epidemiological processes, either by constraining host-to-host transmission or partitioning environmental exposure to microorganisms. Given that social groups can often be identified with limited behavioral data and without molecular data, these findings provide tractable guidance for conservation and infectious disease management strategies.
Keywords: primate microbiome, microbial ecology, social network analysis

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS5

Changes in Genome GC and tRNA Content Influence the Evolution of tRNA Modifying Enzymes in Bacteria

Gaurav Diwan [1] [1,*], Saurabh Mahajan [1], Deepa Agashe [1]

[1]National Centre for Biological Sciences, Bengaluru, India

Abstract: Bacterial genomes show large variation in genome organization and composition, including traits such as tRNA content and codon use that determine translation efficiency. How has the process of translation evolved to deal with this variation? The known tRNA repertoire in all bacteria is sufficient to decode only up to 56 of the 61 sense codons. This discrepancy is solved by enzymes that modify anticodon bases of tRNAs and expand their decoding ability. Thus, tRNA modifying enzymes (MEs) could compensate for the absence of several tRNA molecules. Determining the evolutionary history of MEs will provide insights into whether the tRNA gene content of bacteria is shaped by the evolution of MEs or vice versa.

We mined ~1100 bacterial genomes for all known anticodon MEs and reconstructed the history of enzyme gain and loss on the bacterial phylogeny. We found that most MEs were ancestral to the eubacterial clade, and that ME evolution is dominated by multiple independent losses and gains of major ME pathways. In each case, we observed that the loss of the pathway was associated with the retention of unmodified tRNAs. In addition, ME loss was often associated with shifts in genome GC content and codon use. Thus, MEs evolved early in the eubacterial lineage and were lost repeatedly in conjunction with major shifts in genome GC and tRNA content. These results suggest that changes in genome GC content may have allowed diversification of the genetic code via ME evolution across eubacterial lineages.

Expanded summary*: Bacterial genomes show large variation in genome organization and composition. One widely known difference between genomes is the GC content and therefore codon usage of protein coding genes. Given the variation in codon usage across bacterial genomes, the tRNA pool that recognizes codons is thought to have co-evolved to efficiently decode codons. However, an analysis of the known repertoire of tRNA genes in eubacteria shows that only up to 56 of the 61 sense codons can be decoded. This discrepancy is solved by tRNA modifying enzymes (MEs) that modify anticodon bases and expand the decoding ability of tRNA. For example, the MEs cmoA and cmoB modify Uridine in the anticodon wobble position, allowing U to pair with A, G and U. Thus, tRNA MEs could potentially compensate for the absence of several tRNA molecules, and maintain translation efficiency.

Studies involving tRNA MEs so far have focused on the function of the enzymes and very few studies have determined the evolutionary history of MEs. One such study showed that the evolution of two major ME classes led to the divergence of bacterial and eukaryal tRNA content from archaea (Novoa et al 2012, Cell). Thus, it is clear that MEs play a major role in shaping tRNA content across the three kingdoms of life. However, a detailed and systematic analysis of variation in tRNA content and distribution of MEs within a kingdom is lacking. Determining the evolutionary history of MEs in bacteria will provide insights into whether their tRNA gene content is shaped by the evolution of MEs or vice versa. Additionally, how genome GC content and codon usage influence the evolution of these genomic features remains unexplored.

We mined ~1100 bacterial genomes for all known anticodon MEs using Hidden Markov Model homology searches, and reconstructed the history of enzyme gain and loss on the bacterial phylogeny. We found that most MEs were ancestral to the eubacterial clade, and that ME evolution was dominated by eight major, independent losses and two independent gains of ME pathways. In each of these clades and across ME pathways, we observed that the loss of the enzyme pathway was associated with the retention of unmodified tRNAs (which were otherwise absent in the sister clade that retained the ME). Loss of the ME pathway means that the modified tRNA can no longer recognize multiple codons and that alternate tRNA molecules are required to recognize all codons. Thus, retention of unmodified tRNA could compensate for the loss of MEs. To understand factors that led to the loss of MEs, we also measured changes in other genomic features across the sister clades. We found that ME loss was associated with shifts in genome GC content and codon use in 6 of the 9 analysed sister pairs. For example, Actinobacteria showed the loss of two major ME pathways in conjunction with an increased genome GC content that favored codons recognized by unmodified tRNAs. Thus, we propose that in Actinobacteria and other clades, altered GC bias favored the retention of unmodified tRNAs and ultimately allowed the loss of the ME.

Along with previous studies, our results indicate an intricate link between genomic features such as genome GC content, codon usage, tRNA content and the evolution of tRNA MEs. Since tRNA MEs affect a total of 35 different tRNA species, studying the evolution of MEs is important to understand how the genetic code diversified in bacteria. Diversification of the genetic code led to the availability

of a wider variety of codons that could be used in genomes. Therefore, understanding the above aspects will provide clues about how bacterial genomes diverged from each other and led to speciation events.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-123

A fly in the ointment: compensatory mutations in Drosophila gene deletion strains may confound reproducibility

David Rinker [1,*], Tony Capra [1]

[1]Biological Sciences, Vanderbilt University, Nashville, United States

Abstract: Over the past half-century tens of thousands of mutant *Drosophila melanogaster* lines have been generated, many of which have been subsequently maintained as living stocks in fly repositories and laboratories. While researchers assume that the genomes of such "legacy" stocks remain effectively invariant over time, variability in strain-associated phenotypes has been reported. Here, we hypothesize that phenotypic variability may reflect genomic changes that have accrued in these legacy lines. Specifically, we suggest that the genomes of these strains do not always evolve neutrally and may accrue mutations that compensate for genomic imbalances introduced by the originating mutation.

To test this hypothesis, we applied whole genome sequencing to a pair of long established fly strains that lack a key component involved in olfactory signaling. Genome sequences from these lines, along with sequences of wild type (WT) lines were analyzed to characterize and contrast genome wide patterns of variation. Our analysis revealed that, compared to WT, the genomes of the legacy fly strains contain distinct clusters of fixed variation. Furthermore, those variant dense loci were enriched for genes previously described as relating to neurogenesis, odorant reception, and olfactory-mediated learning. Therefore, not only did the mutant genomes display local densities fixed variation, but the new mutations disproportionately impacted loci functionally or phenotypically related to the line's nominal target gene.
If pervasive among legacy *Drosophila* strains, this effect challenges the assumption that such lines should be expected to remain phenotypically invariant over years of maintenance. This complicates the interpretably and reproducibility of study results.

Expanded summary*: The genomic integrity of model organisms—and by extension their phenotypic integrity—has been assumed to be stable over time. Working under this assumption, researchers may assay mutant lines over many generations of the organism with the expectation that the relevant phenotypic outputs will be consistent. However, it has recently been shown that single gene deletion strains of yeast can display independent but parallel patterns of mutations that are dependent upon the deleted gene of interest (GOI). The provocative implication of this finding is that any induced mutation can precipitate non-random genome-wide alterations that may become manifest over time. Should this phenomenon of "genome rebalancing" affect loci that are also involved in the phenotype related to the GOI, the phenotype could change over time. Such phenotypic instability could confound reproducibility in subsequent studies

Here we extend this question to *Drosophila melanogaster*,a model organism that is a powerful model system for the study of many biological processes relevant to human biology. Once generated, mutant fly lines are generally maintained in perpetuity as living stocks, with the oldest of these "legacy" lines having passed through over 1000 generations. Simultaneously, the genomes of these flies are continually evolving within the context of their unique mutant genotypes. We then forward the hypothesis that the genomes of these legacy strains do not always evolve neutrally and may gradually accrue non-random mutations to compensate for genomic imbalances introduced by the originating mutation.

To test this hypothesis, we applied whole genome sequencing to a pair of long established fly strains that lack a key component involved in olfactory signaling. Genome sequences from these lines, along with sequences of wild type (WT) lines were analyzed to characterize and contrast genome wide patterns of variation. Our analysis revealed that, compared to WT, the genomes of the legacy fly strains contain distinct clusters of fixed variation. Furthermore, those variant dense loci were enriched for genes previously described as relating to neurogenesis, odorant reception, and olfactory-mediated learning.

The different distributions of fixed genetic variation in WT and separately maintained mutant *Drosophila* strains suggest that a given line's originating mutation may alter the selective pressure on associated loci (epistatic selection). If such new mutations disproportionately impact loci that are functionally or phenotypically related to the line's nominal target gene, some legacy mutant fly lines may eventually display alterations to phenotypes initially ascribed to them. This effect may complicate the assumption that a strain may be expected to remain phenotypically invariant over many years of maintenance, thus complicating interpretation and reproducibility.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS14

Predictions of linked (and unlinked) purifying selection as null hypothesis to study diversity across genomes.

Josep Comeron*

Abstract: The use of models that incorporate neutral and deleterious mutations as null hypothesis to investigate the presence of other types of selection such as balancing or positive has been a hallmark of molecular evolution studies for decades. Analyses of nucleotide variation across genomes however have often ignored the consequences of deleterious mutations at linked sites (background selection) when analyzing regions with no clear evidence of reduced or absent recombination. Recent studies in *Drosophila melanogaster* have exposed the major influence of background selection shaping the levels of nucleotide diversity across the entire genome thus supporting the use of predictions of background selection as adequate null hypothesis or baseline genome-wide. These studies benefited from the combination of high-resolution recombination maps and detailed genome annotation that allow predictions to follow the genomic distribution of genes and gene structures. Here I expand such approach, including a more detailed recombination map in *D. melanogaster* and the effects of selection at unlinked sites, to expose signals of balancing and positive selection across genomes of different populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-372

Exploring the relationship between genetics, cranial morphology and geography in human populations.

Carlos Stefano Reyna [1,*], Anna-Sapfo Malaspinas [1], Maria Ávila-Arcos [2], Jody Weissmann [3], Marcia S. Ponce De León [3], Christoph P. E. Zollikofer [3]

[1]Institute of Ecology and Evolution, University of Bern, Bern, Switzerland, [2]Laboratorio Internacional de Investigación sobre el Genoma Humano, UNAM, Juriquilla, Queretaro, Mexico, [3]Anthropological Institute, University of Zurich, Zurich, Switzerland

Abstract: It is believed that anatomically modern humans originated in Africa from where they migrated to the rest of the world. Sequential colonization events would imply that nearby human populations "*look more alike"* than distantly located populations. To reconstruct the history of populations, researchers have relied on different types of data, including morphological and genetic data. However, in recent times, the usefulness of cranial morphology to reconstruct human population history has been questioned due to the potential influence of the environment in shaping morphological traits. Recently, it has been shown that cranial morphology can reflect genetic distances across populations.

The unprecedented increase in the number of available whole genomes makes it now possible to reinvestigate the relationship between traditionally collected morphological traits with large scale genomic datasets. In this work, we assemble large public genomic datasets from worldwide human populations (1000G, HGDP and SGDP) and compare these data with phenotypic data of geographically matching populations, as represented by the well-known Craniometrics data set of W.W. Howells. We study the relationship between these two types of datasets and the dispersal distance from Africa for the sampled populations and attempt to highlight some factors that impact the degree of correlation among datasets.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-338

Signatures of multiple-mergers coalescence in genomic diversity data

Daniel Rice [1,*], John Novembre [1], Michael Desai [2]

[1]Human Genetics, University of Chicago, Chicago, [2]Organismic and Evolutionary Biology, Harvard University, Cambridge, United States

Abstract: The genetic diversity of a population reflects its demographic and evolutionary history. Methods for inferring this history typically assume that the ancestry of a sample can be modeled by the Kingman coalescent process. A defining feature of the Kingman coalescent is that it generates genealogies that are binary trees: no more than two ancestral lineages may coalesce at the same time. However, this assumption breaks down under several scenarios. For example, pervasive natural selection, rapid spatial range expansion, and extreme variation in offspring number can all generate genealogies with "multiple-merger" events in which more than two lineages coalesce instantaneously. Therefore, detecting multiple mergers is important both for understanding which forces have shaped the diversity of a population and for avoiding fitting misspecified models to data. Current methods to detect multiple mergers rely on the average site frequency spectrum (SFS). However, many of the signatures of multiple mergers in the average SFS are also consistent with a Kingman coalescent process with a time-varying population size. Here, I present a new method for detecting multiple mergers based on correlations in the SFS across the genome. Unlike the average SFS, these correlations depend mostly on the topologies of genealogies rather than their branch lengths and are therefore robust to most demographic effects. We apply this method to genomic diversity data from a variety of species.

Expanded summary*: The goal of population genetics is to understand patterns of genetic diversity in terms of evolutionary and demographic factors such as the historical population size, rates of migration, and the influence of natural selection. To this end, researchers use simplified models of reproducing populations, which trade a degree of accuracy for enhanced generality and mathematical tractability. A fundamental problem for population genetics is that genetic diversity is determined by different factors in different biological scenarios, so no single class of models is appropriate for all situations. Therefore, it is crucial to be able to identify when a particular model is misspecified and select a more appropriate one. Not only is this model-checking process a necessary step in validating inferences, but also it can yield important information about the dominant evolutionary forces acting on a population.

The most commonly-used family of models, known as coalescent models, use the genealogical history of the sample as an intermediate step in deriving patterns of diversity from population histories. While there exists a variety of coalescent models derived from different biological assumptions, almost all population genetic inference methods are based on the Kingman coalescent Under this model, the average level of genetic diversity in a population reflects its historical size. Elaborations of the Kingman coalescent include time-dependent population size and migration among isolated subpopulations. Thus, the Kingman coalescent is useful for inferring the demographic history of a population.

While the Kingman coalescent is robust to some perturbations in the underlying population model, there are a number of biologically relevant scenarios in which it does not apply. One particularly interesting violation of the Kingman assumptions is natural selection. If a region of the genome contains a sufficient number of sites under selection, sample genealogies will not follow a Kingman distribution, and genetic diversity will not (directly) reflect the population size, rendering traditional population genetic inference methods invalid. It is therefore important to be able to detect violations of the Kingman coalescent, especially as population genetics expands to genome-scale studies of a larger number of species that vary more widely in their evolutionary and demographic histories.

In my postdoctoral research, I am developing methods for distinguishing among coalescent models and apply these methods to population genomic data for a variety of species. This work serves two purposes. First, by identifying violations of the Kingman coalescent, it will allow population geneticists to use standard inference methods only where appropriate and thus feel more confident in the results. Second, it will help us understand the relative influence of various evolutionary forces across genomic regions and species.

I have found a promising candidate data summary based on the site frequency spectrum (SFS). I have found, via simulation, that multiple mergers induce local correlations between the numbers of mutations at similar frequencies. For example, a genomic region containing a mutation at high frequency in a sample is likely to contain an above-average number of other high-frequency mutations.

These correlations are negative in simulations of the Kingman coalescent and this finding appears to be robust to demographic fluctuations.
I am applying this method to test the Kingman assumptions in genomic data from natural populations. In a preliminary study, I have calculated the local correlations in the SFS to genomic data from a large sample of Drosophila melanogaster. My initial results suggest that the Kingman coalescent is not a good model for this data, suggesting that natural selection may be pervasive in the genome and calling into question standard demographic inference methods in Drosophila. I am currently exploring alternative explanations for the correlations in the SFS in this data set that would be consistent with the Kingman model. Beyond this Drosophila data, I intend to calculate the SFS correlations to the large cross-species genomic data set compiled recently by Corbett-Detig and Sackton.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-329

Mitochondrial DNA reveals cultural and demographic influences on Native American population history in the southern United States

Aida Miro-Herrans [1,*], Marcus Briggs-Cloud [2 3], Ana Sylestine [4], Deborah A Bolnick [1 5]

[1]Department of Anthropology, University of Texas at Austin, Austin, [2]Maskoke, [3]School of Natural Resources and the Environment, University of Florida, Gainesville, [4]Coushatta Tribe of Louisiana, Elton, [5]Population Research Center, University of Texas at Austin, Austin, United States

Abstract: Relatively little attention has been given to recent evolutionary history and the impact that European contact has had on Native American genetic diversity over the last five centuries. Since European contact, Native Americans have experienced declines in population size, forced migrations, community reorganization, and genetic exchange with non-native peoples. The southern United States represents one of the earliest sites of European contact with Native Americans, yet little is known about how the demographic history in the region may have impacted the genetic patterns of the population. In this study, we investigated how indigenous mtDNA diversity in the southern U.S. has been shaped by matrilineal clan affiliations and demographic history. This project was developed in collaboration with members of participating communities, and it was designed to address questions important to research participants. We collected saliva samples from 80 Native Americans in the southern U.S. and sequenced whole mtDNA genomes. We analyzed haplotype diversity patterns together with cultural and demographic data to evaluate the factors shaping mtDNA diversity in this region. We tested for changes in population size since European contact. We also performed simulations to test specific hypotheses to explain current genetic diversity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-336

Base composition convergence misleads species tree inference: a case study from the malaria parasites (order Haemosporida)

Spencer Galen [1,*], Susan Perkins [1]

[1]Invertebrate Zoology, American Museum of Natural History, New York, United States

Abstract: The recent growth in the size of phylogenomic datasets has made it critical to recognize and account for non-phylogenetic signal that has the potential to obscure evolutionary history. One source of non-phylogenetic signal, base composition bias, is highly dynamic across the tree of life and can lead to sequence convergence in distantly related taxa. However, the effect of base composition bias on modern phylogenomic methods has been underexplored. Here we investigate the effect of convergence in base composition and codon usage profiles on supermatrix and species tree inference using a phylogenomic dataset from the malaria parasites (order Haemosporida). We find that both supermatrix and species tree methods are mislead by "base composition attraction" between the outgroup *Theileria annulata* and a clade of primate-infecting *Plasmodium* parasites that have secondarily evolved elevated GC content. Without correcting for base composition both methods recovered a topology in which the GC-rich *Plasmodium* clade is driven towards the outgroup as sister to the rest of the malaria parasites; however, when base composition is corrected for, we recovered a highly divergent topology in which primate *Plasmodium* is found in a derived position and the avian malaria parasite genus *Leucocytozoon* is found at the base of the tree. This study has significant implications for our understanding of life history evolution in the malaria parasites, and demonstrates the importance of accounting for non-phylogenetic signal in phylogenomic studies when base composition is heterogeneous.

Expanded summary*: My research contributes to advancements in two distinct areas of study in systematics and molecular evolution: 1) the evolutionary history of a clade of virulent blood pathogens, the malaria parasites, and 2) the effects of base composition bias on phylogenomic inference. First, this study represents the most comprehensive dataset constructed for any malaria phylogeny to date, and includes several poorly studied and enigmatic malaria genera that provide critical context to our understanding of the evolution of the human malaria parasites. Previous studies of malaria phylogeny have been hampered by restricted taxon sampling and a limited number of molecular markers, which has resulted in controversy regarding the pattern and process of diversification across the major lineages of the malaria phylogeny. Using a 22,000 bp protein-coding gene dataset, we recovered a novel topology that confirms the "bird first" hypothesis of malaria evolution that suggests an avian origin for the malaria parasites followed by a single invasion of mammals as hosts. The novel topology that I will present also provides important insights into the evolution of malaria life history traits, timing of diversification, and the history of host-switching in this globally important clade of pathogens. Second, this research also represents a case study for the effect of base composition bias and heterogeneity on supermatrix and species tree phylogenetic estimation, a problem that is likely to increase in frequency as phylogenomic datasets continue to expand in size. Understanding the effects of base composition bias on phylogenomic inference is of particular importance because base composition is not only heterogeneous across the tree of life, but can also vary between closely related species within the same clade. It is possible to mislead phylogenetic estimation when distantly related taxa evolve convergent base composition, though this effect has been poorly explored with large phylogenomic datasets. The malaria parasites represent an ideal group with which to explore the effect of base composition convergence, as mean GC content varies by as much as ~25% among taxa within the same genus. My research shows that both supermatrix and species tree approaches are mislead by base composition convergence in distantly related taxa, though this effect is erased when base composition heterogeneity is corrected for. Importantly, the conflicting topologies posit dramatically different scenarios for trait evolution and the pattern of host-switching in the malaria parasites: standard analyses recover primates as the ancestral host to the malaria parasites, while corrected analyses recover birds at this position. This finding has wide-ranging implications for systematic studies of taxonomic groups that are characterized by heterogeneous base composition. Furthermore, this research will provide an example of the importance of interrogating phylogenomic data for base composition heterogeneity in poorly studied clades for which this trait may not yet be characterized. In sum, my research provides a critical advancement to our understanding of malaria parasite systematics and trait evolution, as well as a valuable case study of the effect of base composition heterogeneity on phylogenomic inference.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OT-OS6

Biased gene conversion drives codon usage in human and precludes selection on translation efficiency

Fanny Pouyet [1 2,*], Dominique Mouchiroud [1], Laurent Duret [1], Marie Sémon [3]

[1]Laboratoire de Biometrie et biologie evolutive, CNRS - Univeristy Lyon 1 - University of Lyon, Villeurbanne, France, [2]Biology, Institute Ecology and Evolution, Bern, Switzerland, [3]Laboratoire de Biologie et modelisation de la cellule, ENS Lyon - University of Lyon, Lyon, France

Abstract: In humans, as in other mammals, synonymous codon usage (SCU) varies widely among genes. In particular, genes involved in cell differentiation or in proliferation display a distinct codon usage, suggesting that SCU is adaptively constrained to optimize translation efficiency in distinct cellular states. However, in mammals, SCU is known to correlate with large-scale fluctuations of GC-content along chromosomes, caused by meiotic recombination, via the non-adaptive process of GC-biased gene conversion (gBGC). To disentangle and to quantify the different factors driving SCU in humans, we analyzed the relationships between functional categories, base composition, recombination, and gene expression. We first demonstrate that SCU is predominantly driven by large-scale variation in GC-content and is not linked to constraints on tRNA abundance, which excludes an effect of translational selection. In agreement with the gBGC model, we show that differences in SCU among functional categories are explained by variation in intragenic recombination rate, which, in turn, is strongly negatively correlated to gene expression levels during meiosis. Our results indicate that variation in SCU among functional categories (including variation associated to differentiation or proliferation) result from differences in levels of meiotic transcription, which interferes with the formation of crossovers and thereby affects gBGC intensity within genes. Overall, the gBGC model explains 70% of the variance in SCU among genes. We argue that the strong heterogeneity of SCU induced by gBGC in mammalian genomes precludes any optimization of the tRNA pool to the demand in codon usage.

Expanded summary*: There is a recurrent debate in the scientific community about the processes (selective or not) that drive patterns of synonymous codon use (SCU) in the human genome. Both adaptive and non-adaptive processes, which are not mutually exclusive, have been proposed to explain the existence of codon usage biases. According to the main adaptive model, termed translational selection, SCU and abundance of tRNA are co-adapted to optimize the efficiency of translation. Non-adaptive models propose instead that codon usage bias results from biases in neutral substitution patterns, driven by mutation or by GC-biased gene conversion (gBGC). Regularly, new papers are published pretending to demonstrate that human genes are subject to translational selection but without considering the possibility this could also be a consequence of non-adaptive processes. Here, we examine why SCU differs between two major functional classes of genes, i.e. proliferating versus differentiating genes, as it has previously been suggested that this is a consequence of selection upon translation. We demonstrate that SCU variation cannot be explained by such a process and it is in fact pefectly consistent with an effect of gBGC, as follows: we first replicate the pattern previously discovered that proliferating and differentiating genes have different SCU. However, we show that this can be summarised as a difference in GC content. We show this is not associated with translation because the same patterns are found for amino acids matched by a single tRNA and those matched by multiple tRNAs.

Our research is not simply a rebuttal of translational selection: we then show the difference in GC content can be explained by different levels of intragenic recombination rate, and the difference in recombination rate can be explained in terms of differences in expression. We show that the level of gene expression in meiotic cells affects substitution patterns and hence SCU (because transcription affects meiotic recombination, and hence gBGC). We quantify the influence of gBGC at 70% of SCU variation. Finally, we argue that the strong heterogeneity of SCU induced by gBGC in mammalian genomes precludes any optimization of the tRNA pool to the demand in codon usage.

There is a strong interest in identifying the genomic features that contribute to the proper functioning of organisms in our scientific community. In this neat analysis, I accurately prove the variation of synonymous codon use in humans is not due to translational selection but to gene conversion during recombination. This analysis is broadly significant as I demonstrate the importance of taking

simultaneously into account these two hypotheses to distinguish between them, and that is why this research perfectly matches the Fitch symposium.

Keywords: Synonymous codon usage, translational selection, biased gene conversion

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-406

A full-likelihood coalescent method or detecting sites under selection

Aaron Stern [1,*], Rasmus Nielsen [2]

[1]Graduate Group in Computational Biolo, [2]Integrative Biology, UC Berkeley, Berkeley, United States

Abstract: A major goal of genetics research is to identify selection throughout the genome and understand how selection shapes neighboring genetic diversity via selective sweeps. While there exist numerous methods to detect signatures of selective sweeps from sequence data, these methods are limited in several ways: conflation of selection with factors such as demography and restricted power outside of detecting the particular type of sweep to which an individual method is tuned. To this end, we propose a novel method for estimating the likelihood ratio (LR) that a site has evolved under selection vs neutrality. This method is the first full-likelihood test for selection; i.e., the test makes full use of the latent ancestral recombination graph (ARG) that summarizes all recombination and coalescent events within the sample. We perform importance sampling on ARGs to obtain a consistent estimate of the LR at a particular site. This importance sampling approach also allows us to save great computational expenses in using the estimated LRs to perform hypothesis testing for selection on a genomewide level. In simulations, we show our method has improved power over state-of-the-art sweep detection methods across diverse types of sweeps, and we also apply our method to a scan for recent selection using 54 human genomes from Complete Genomics.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS11

Expression evolution of Drosophila nested genes

Raquel Assis*

Abstract: Nested genes are the most common form of protein-coding overlap in eukaryotic genomes. Previous studies have shown that nested genes accumulate rapidly over evolutionary time, typically via the insertion of short young duplicate genes into long introns. However, the evolutionary relationship between nested genes remains unclear. Here, I compare RNA-seq expression profiles of nested, proximal intra-chromosomal, intermediate intra-chromosomal, distant intra-chromosomal, and inter-chromosomal gene pairs in two *Drosophila* species. I find that expression profiles of nested genes are more divergent than those of any other class of genes, supporting the hypothesis that concurrent expression of nested genes is deleterious due to transcriptional interference. Further analysis reveals that expression profiles of derived nested genes are more divergent than those of their ancestral un-nested orthologs, which are more divergent than those of un-nested genes with similar genomic features. Thus, gene expression divergence between nested genes is likely caused by selection against nesting of genes with insufficiently divergent expression profiles, as well as by continued expression divergence after nesting. Moreover, expression divergence and sequence evolutionary rates are elevated in young nested genes and reduced in old nested genes, indicating that a burst of rapid evolution occurs after nesting. Together, these findings suggest that similarity between expression profiles of nested genes is deleterious due to transcriptional interference, and that natural selection addresses this problem both by eradicating highly deleterious nestings and by enabling rapid expression divergence of surviving nested genes, thereby quickly limiting or abolishing transcriptional interference.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OT-OS2

A very recent whole genome duplication in Potamopyrgus antipodarum predates multiple origins of asexuality & associated polyploidy

John Logsdon [1,*], Maurine Neiman [1], Jeffrey Boore [2], Joel Sharbrough [3], Laura Bankers [1], Kyle McElroy [1], Joe Jalinsky [1], Peter Fields [4], Peter Wilton [2]

[1]Department of Biology, University of Iowa, Iowa City, [2]Department of Integrative Biology, University of California, Berkeley, [3]Department of Biology, Colorado State University, Fort Collins, United States, [4]Zoologisches Institut, Universität Basel, Basel, Switzerland

Abstract: *Potamopyrgus antipodarum*, a New Zealand freshwater snail, is a powerful system to study the maintenance of sexual reproduction. Obligate asexual *P. antipodarum* lineages include both triploids and tetraploids that are products of multiple separate transitions from diploid sexual ancestors. Distinct diploid sexual and polyploid asexual lineages coexist and compete; these separate lineages can be considered replicated natural experiments. We have shown that harmful mutations are accumulating at a higher rate in asexual than in sexual *P. antipodarum*, demonstrating the utility of this system as a model for investigating the evolution of sex at the genomic level. In order to better understand the causes and consequences of transitions to asexuality, we have sequenced multiple genomes and transcriptomes of *P. antipodarum* and a close relative, *P. estuarinus*, a diploid sexual species. The diploid genome size of *P. estuarinus* is ~0.6X of the genome size of diploid *P. antipodarum*, inspiring us to investigate whether the most recent common ancestor of *P. antipodarum* had experienced a whole-genome duplication (WGD) event prior to the diversification of its many sexual and asexual lineages. In addition to its clear relevance to understanding the evolutionary history of this species, by being so recent, this apparent WGD will also be especially powerful in understanding events immediately following WGD. Our initial genome assembly of a model sexual *P. antipodarum* lineage was consistent with this possibility, indicating high fractions (~35%) of scaffolds containing extended, nearly identical, duplicated regions. This result also partly explains our general difficulty with assembling the genome, despite generating >100X genome coverage using multiple methodologies. Even considering the limitations of our current genome assembly, we used the assembly to test a series of predictions under the hypothesis of recent whole-genome duplication, all of which are consistent with WGD. These include: 1) a marked excess of duplicated copies of genes in *P. antipodarum* which are maintained in single copy in other animals, 2) implausibly high "heterozygosity" estimates in our model *P. antipodarum* sexual genome, presumably resulting from non-allelic comparisons, 3) higher sequence identity between thousands of *P. antipodarum*-specific paralogous genes, when compared to their *P. estuarinus* orthologs. These and additional lines of evidence will be presented and evaluated. Together, our results hint that this initial genome-wide duplication event might have played a key role in the subsequent evolutionary trajectory of this species, potentially facilitating its repeated diversification into multiple asexual lineages. We are now generating additional long-range genome scaffolds for *P. antipodarum* using multiple methods, as well as improving the coverage and quality of the *P. estuarinus* genome. We will use these new data to conduct definitive phylogenomic tests of this especially remarkable whole genome duplication.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-385

De novo assemblies and comparative genomics of Streptomyces aureofaciens strain ATCC 10762 and related species

Etsuko Moriyama [1,*], Julien Gradnigo [1], Greg Somerville [2], Michael Huether [3]

[1]School of Biological Sciences, [2]School of Veterinary Medicine and Biomedical Sciences, University of Nebraska-Lincoln, [3]Zoetis, Lincoln, United States

Abstract: *Streptomyces aureofaciens* is a Gram-positive Actinomycete used for commercial antibiotic production. Although it has been the subject of many biochemical studies, no public genome resource was available until very recently and phylogenetic placement and genomic contents were unknown. To address this need, the genome of *S. aureofaciens* strain ATCC 10762 was sequenced by Illumina MiSeq and Roche 454 pyrosequencing. The genome assemblies generated by multiple *de novo* assembly methods (*i.e.*, IDBA-UD, MIRA, SGA, SOAPdenovo2, SPAdes, and Velvet) were assessed. Empirical sequence data from targeted PCR of predicted gap regions provided a validation framework for the assemblies. Overall, the best assembly was generated using SPAdes with both Illumina and 454 reads. The total length of the final assembly was 9.24 Mb and the average G+C content was 72.7%, both in line with what expected before. This assembly was annotated using the NCBI Prokaryotic Genome Annotation Pipeline, revealing a total of 8,076 genes. Functions were predicted for approximately 60% of the 7,630 protein-coding genes. With additional analysis, the putative functions of another 10% of these protein-coding genes were predicted. Comparative genomics, phylogenomic analysis, and gene content demonstrate that *S. aureofaciens* strain ATCC 10762 is closely related to the genus *Kitasatospora*, forming an evolutionarily distinct group separate from the *Streptacidiphilus* and *Streptomyces* genera within the *Streptomycetaceae* family. We will also discuss the effect of using long-read sequencing data (Nanopore) in addition to short-read data (Illumina and 454) in de novo assembly.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OT-OS5

Tracing the origins of 19th century enslaved Africans

Marcela Sandoval Velasco [1,*], Victor Moreno-Mayar [1], Maria Ávila-Arcos [2], Kate Robson-Brown [3], M. Thomas P. Gilbert [1], Hannes Schroeder [1]

[1]Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, [2]International Laboratory for Human Genome Research, National Autonomous University of Mexico, Queretaro, Mexico, [3]Department of Archaeology, University of Bristol, Bristol, United Kingdom

Abstract: The transatlantic slave trade has been the largest movement of people in recorded history. Between 1500 and the 1900, over 12 million Africans were enslaved and forcedly transported to the Americas and islands in the Atlantic Ocean. The island of Saint Helena in the South Atlantic played an important role during abolition times when the British Royal Navy was intercepting slave ships and disembarking the cargo in the island. During the second half of the 19th century, around 26,000 Africans were disembarked on Saint Helena. Historical records suggest that these people had originated from West, West Central Africa and East Africa, but their precise origins remain unknown. We generated low-coverage genome-wide data from enriched ancient DNA libraries from over 50 enslaved individuals. We examine their geographic origins by comparing them to the increasingly available reference panels of modern genomic variation in Africa. Additionally to exploring their genomic ancestry, we explore relatedness and admixture patterns, in order to assess population diversity during the slave trade. By tracing geographic source populations to different sub-continental modern populations within Africa we attempt to shed new light on the origins of enslaved Africans and the workings of the transatlantic slave trade during that time period.

Expanded summary*: From 1840 to 1870, during the last stage of the transatlantic slave trade, the remote South Atlantic British island colony of Saint Helena played an important role in the suppression of the slave trade years after the British had established abolition in 1807. During the early years of the transatlantic slave trade, Saint Helena played only a secondary role, mainly acting as staging and resupply post for warships cruising south of the equator returning from Africa. By the 1840s Saint Helena started serving as provisioning station and stopping point for ships returning from the East Indies. England's Vice-Admiralty Court started using Saint Helena as a trial venue and receiving depot for slave vessels captured at sea. From 1839 to 1867 a total of 26,000 enslaved Africans came to Saint Helena, of which 8,000 are thought to be buried on the island, while around 18,000 Africans survived to be liberated by Saint Helena's court. What happened to the enslaved after 'liberation' is unknown in most individual cases, but the historical record suggest that the majority emigrated to Cape Colony (todays Cape Town) and to the West Indian Colonies of Jamaica, Trinidad and Guiana.

Although the approximate source region of the enslaved Africans is known, more specific locations are uncertain, as are whether the enslaved populations originated from tight and geographically limited areas, or whether they originated from distinct geographical locations of Africa and were forced to travel long distances to the coastal regions of embarkation points. Also, it is unknown if the dynamics and extent of the trade varied over time.

Saint Helena offers a unique insight into the dynamics of the last phase of the transatlantic slave trade, in particular, the randomly sampled skeletons should offer an insight into the range of places where people were coming from when captured during the last few decades of the slave trade. To investigate this and in an attempt to shed light on the origins and ancestry of these people, we analyzed ancient DNA genome-wide data from human archaeological remains of enslaved individuals from the island of Saint Helena. Genome-wide analyses of single nucleotide polymorphisms provide a powerful tool for estimating individual ancestry, and several studies have shown that they can be used to infer an individual's geographic origin with surprising accuracy. Using low-coverage ancient DNA genome wide data, we investigate relatedness, admixture patterns and the geographic origins of the enslaved individuals taken to the island between 1840 and 1870, aiming to deepen our understanding of the extent of the trade inside Africa and its dynamics during abolition times on the second half of the 19th century.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-390

Determining the Relationship Between T Cell Activity and Affinity of Public vs. Private Cytomegalovirus-specific TCRs

Christopher Stevens [1,*], Chad Williams [2], Zach Frye [1], Jenny Jiang [2], Jennifer Maynard [1]

[1]Chemical Engineering, [2]Biomedical Engineering, University of Texas at Austin, Austin, United States

Abstract: Human cytomegalovirus (hCMV) infections elicit T cell responses that are sometimes dominated by $CD8^+$ cytotoxic T lymphocytes (CTLs) bearing highly related or identical public TCRs in unrelated individuals. Due to the persistence of CMV infections it evolves to fit its host suggesting that there may be some evolutionary bias towards selection of these public TCRs. To understand the driving force behind this selection we sought to identify a correlation between the affinity and the activity of several public and private TCRs specific to HLA-A*0201-restricted CMV antigen ($pp65_{495-503}$) isolated from a diverse population of CMV+ patients. Previously, we have developed a robust system for rapid screening of TCR binding and activation using soluble and membrane bound DO11.10 TCR. We applied this system to the immunodominant RA14 TCR by fusing the variable regions to mouse constant domains and transfecting into bald $58^{-/-}$ T cell hybridoma. IL-2 secretion in the presence of immobilized pp65/HLA tetramer revealed specific activation of the chimeric RA14 transfectants. Affinity was determined using a cellular interaction approach where individual transfectants were introduced to red blood cells functionalized with pMHC to measure the TCR/pMHC interaction *in situ*. This '2D affinity' for RA14 was found to be consistent with measurements by SPR and for similar affinity TCRs. We have shown that the higher affinity public pp65-specific TCRs secreted significantly higher amounts of IL-2 when exposed to immobilized pp65/HLA tetramer compared to the lower affinity private TCRs emphasizing a physiological selection towards public TCRs to elicit a more potent T cell response.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OM-OS1

Mitonuclear transcriptomes uncover common features of epistasis (GxG) and genotype x environment (GxE) interactions in Drosophila

David Rand [1,*], James Mossman [1], Zhijin Wu [2]

[1]Ecology & Evolutionary Biology, [2]Biostatistics, Brown University, Providence, United States

Abstract: Epistasis (GxG) and genotype by environment interactions (GxE) are usually considered distinct genetic phenomena. But they share the common feature that some factor external to a focal gene modifies the phenotypic effects of that gene. These interaction effects may contribute to the missing heritability of complex traits. We explore the possibility that some genes contribute to both GxG and GxE effects on transcriptional variation, using a subset of 72 mitonuclear genotypes constructed from all pairwise combinations of 6 mtDNAs of *D. melanogaster* and *D. simulans* and 12 of the Drosophila Genetic Reference Panel (DGRP) strains. These genotypes show considerable mitonuclear epistasis for development time, which is modified by high vs. low protein diets (Mossman et al. 2016, Genetics). We have quantified transcriptional variation in both males and females of 12 of these mitonuclear genotypes, each exposed to high vs. low-protein diets, and sampled at four time points. This design, which includes 192 RNAseq libraries, allows us to partition variation among four primary factors: nuclear genes, mtDNA genes, sex and diet, plus pairwise and higher order interaction effects that can quantify GxG and GxE. We test the hypothesis that genes exhibiting mitonuclear GxG effects are independent of genes contributing to mitonuclear GxE effects, using analyses of variance for each transcript. We show that there is significant overlap in the genes attributed to these effects, consistent with the alternative hypothesis that GxG and GxE share common features in quantitative genetic variation related to mitochondrial and metabolic function. However, each mitonuclear genotype has a unique transcriptional response to alternative diets. These data point to the complexity of personalized genomic medicine, but also suggest that there are common elements to epistasis and genotype-by-environment interactions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-359

Insight into the plastomic evolution in Podocarpaceae (Gymnosperms)

Edi Sudianto [1 2 3,*], Chung-Shien Wu [3], Lars Leonhard [4], William F. Martin [5], Shu-Miaw Chaw [2 3]

[1]Department of Life Science, National Taiwan Normal University, [2]Biodiversity Program, Taiwan International Graduate Program, [3]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan, [4]Botanical Garden, [5]Institute of Molecular Evolution, Heinrich Heine University, Düsseldorf, Germany

Abstract: The plastid genomes (plastomes) of cupressophytes are characterized by loss of canonical inverted repeats (IRs), frequent recombination, genome compaction and reduction, and emergence of novel short IRs. However, previous studies on the cupressophyte plastomes were heavily biased towards sampling of Cupressaceae. Podocarpaceae is the second largest family in cupressophytes with 156 species distributed mainly in Southern Hemispere. To date, plastomes from only four Podocarpaceous species are available in GenBank. They are restricted to two sister clades, the podocarpoids and the dacrydioids. We determined plastome sequences from three phylogenetically distant Podocarpaceae species: *Phyllocladus aspleniifolius*, *Lagarostrobos franklinii*, and *Microstrobos fitzgeraldii*. The plastome of *Lagarostrobos* (151,567 bp) is the largest among cupressophyte plastomes (121–146 kbp) studied so far, suggesting that plastomic expansion likely occurred after the divergence of *Lagarostrobos* from other Podocarpaceae. We also found that the plastome of *Lagarostrobos* contains ~16 pseudogenes, which contribute to its relatively high content of non-genic sequences (43.5%) as compared to that of other Podocarpaceous plastomes (~38.0%). In addition, reconstruction of plastomic inversions across the evolution of Araucariales (including Podocarpaceae and Araucariaceae) revealed that the absolute rate of inversions ranges from 0 to 6.38 times per 100 MYA among Podocarpaceous genera. Disparity in the inversion rate implies that lineage effects rather than evolutionary time drive the plastomic rearrangements in Podocarpaceae. These new data provide new perspectives on plastome evolutionary dynamics in Podocarpaceae and cupressophytes.

Expanded summary*: The cupressophytes (or the conifers II) is one of the most diverse groups in gymnosperms. They comprise ca. 400 species across five families: Cupressaceae, Taxaceae, Sciadopityaceae, Podocarpaceae, and Araucariaceae. The Cupressaceae and Podocarpaceae jointly compose about 75–80% of the cupressophytes species. Early studies showed unique characteristics in the plastid genomes (plastomes) of this group. The cupressophytes plastomes have lost their canonical inverted repeats (IRs) and undergone frequent plastomic recombination. More recently, large-scale analyses showed that the cupressophytes plastomes evolve towards genome compaction and reduction. However, these studies sampling were mainly heavily biased towards the Cupressaceous species. To date, only four Podocarpaceous plastomes are available in the GenBank. The four available species are closely related genera and mainly distributed in the Northern Hemisphere. To provide a better picture of the plastomic evolution in Podocarpaceae, we deciphered three Podocarpaceous genera of the Southern Hemisphere: *Phyllocladus aspleniifolius*, *Lagarostrobos franklinii*, and *Microstrobos fitzgeraldii*.

We found that the *Lagarostrobos* plastome (151,567) is the largest of the three or any other elucidated cupressophytes (121–146 kb). This suggest that the *Lagarostrobos* plastome has likely experienced a genome expansion, contradicting general trend of genome compaction and reduction in cupressophytes. The expansion of *Lagarostrobos* plastome is contributed by proliferation of pseudogenized copies of plastid genes, including *clpP, infA, ndhJ,* etc. We found approximately 16 pseudogenes distributed across the *Lagarostrobos* plastome. These have increased the non-genic sequences content of the *Lagarostrobos* (43.5%) as opposed to that of other Podocarpaceae (~38.0%). On the contrary, the plastomes of *Phyllocladus* and *Microstrobos* are only 2–3 kb larger than the elucidated Podocarpaceous plastomes, and no significant differences were observed in their non-coding content. The tRNA repertoire of *Lagarostrobos* plastome also shows proof of duplications with presence of three copies of two tRNAs (*trnP-GGG* and *trnN-GUU*) and two copies of *trnL-CAA*. Except for *trnN-GUU*, these tRNAs are only found as single-copy in the plastomes of other Podocarpaceae and cupressophytes.

In addition, we reconstructed the plastomic inversions history of Araucariales (including Podocarpaceae and Araucariaceae) to infer the inversion events in Podocarpaceae. Our estimates of absolute rate of plastomic inversions in Podocarpaceae ranges from 0 to 6.38 times per 100 MYA. The observed variations among Podocarpaceous plastomes inversion rates support the notion that plastomic inversions are not influenced by their evolutionary time, but rather due to lineage effects.

In summary, these newly acquired data would supplement the current pool of intriguing plastomic variation of Podocarpaceae and cupressophytes. In particular, the *Lagarostrobos* plastome provide a unique case where it goes against the prevailing trends of plastomic compaction and reduction in cupressophytes. The results demonstrate that there is more to cupressophytes plastomes than meets the eye. The frequent inversions contradicts previous studies that show conserved plastid operons in the Podocarpaceae. It would be interesting to further investigate the influence of these inversions to the efficiency of plastid operon transcriptions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-380

The effects of insulin signaling on sexually dimorphic gene expression in Drosophila

Rita Graze [1,*], RueiYing Tzeng [2], Tiffany Howard [1], Michelle Arbeitman [2]

[1]Biological Sciences, Auburn University, Auburn, [2]College of Medicine, Florida State University, Tallahassee, United States

Abstract: Insulin signaling functions in nutrient sensing and growth. The core IIS/TOR pathway is highly conserved, from fruit flies to man. In many organisms insulin signaling plays a role in regulation and development of sexually dimorphic traits, including body size dimorphism, activity level dimorphism and in female fertility and mating behaviors. To understand how the insulin signaling pathway contributes to sexually dimorphic gene expression, we examined similarities and differences in the effect of perturbation of the pathway on gene expression in male and female Drosophila. Expression of a dominant negative *InR* transgene (InRDN) was driven in the adult stages only by a drug inducible, ubiquitously expressed, "GeneSwitch" GAL4 system. Expression was assessed by RNA-seq of head tissues, in replicate for each sex expressing InRDN, and for genetically matched controls. Males and females have a shared regulatory response to the perturbation, which as expected is heavily enriched for genes and pathways involved in metabolism. However, there are a large number of genes which show striking sex differences only under the perturbation conditions; conditions which decrease insulin signaling, mimicking a starvation-like stress. Perhaps surprisingly, this includes large effect changes in expression of immune, defense and stress response genes driven by male-specific effects of the perturbation. Finally, a subset of genes are dimorphically expressed only when insulin signaling functions normally. These include energy homeostasis genes regulated by insulin signaling, including those known to be dimorphically expressed in Drosophila, for example *sxe2*. Collectively our results suggest that insulin signaling is important for sex differences in energy homeostasis and may also mediate differences between males and females in the relative response to some sources of stress, which has broad implications for physiological underpinnings of trade-offs, sexual conflict and Bateman's principle.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-410

Intra-specific variation in recombination rates as explanation for differences in levels of diversity among D. melanogaster populations

Johnny Cruz Corchado [1,*], Josep M Comeron [1 2]

[1]Interdisciplinary PhD Program in Genetics, [2]Department of Biology, University of Iowa, Iowa City, United States

Abstract: Recombination is a crucial biological process that also plays a key role in evolution. Yet recombination is itself an evolving trait that varies between closely related species and among populations (and individuals) of the same species. In our study, we estimated recombination rates in five *Drosophila melanogaster* populations (Zambia, Rwanda, Cameroon, France and USA) based on population genomics data and determined the inter-population variation in recombination rates across the genome. We observe that recombination rates not only change in total magnitude but also, and significantly, in their relative distribution within chromosomes (recombination landscapes). We then analyze if the observed differences in recombination landscapes play a significant role explaining population-specific differences in nucleotide diversity under a linked selection scenario that considers only purifying selection (i.e., Background Selection). Our results suggest that population-specific differences in nucleotide diversity at specific genomic regions can be explained by differences in population-specific recombination rates and the corresponding linked selection effects associated with the inevitable and frequent input of deleterious mutations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-383

Understanding genetic disease in the context of evolutionary history, genome duplication and the evolution of function

Alexandra Martin-Geary*, Mark Reardon [1], Nikita Abramovs [1], May Tassabehji [1], David Robertson [1]

[1]University Of Manchester, Manchester, United Kingdom

Abstract: Our understanding of heritable variation is heavily informed by what we know about the processes and biological factors associated with their genomic context. This is inclusive of evolutionary histories such as paralog status, evolutionary age, and the broader downstream properties of a gene; such as its haplosufficiency, number of protein interactions, functional constraints and propensity to accumulate variants. Using various statistical methods, such as principle component analysis, on a large dataset generated from open-source biological data pertaining to diverse genetic characteristics, we show that there are a number of factors which are peculiar to genes associated with dominant and recessive disorders, which are at odds with patterns observed in genes with no known disease association. We discuss how these factors impact on our current understanding of the role of both evolutionary and downstream factors contributing to a gene's propensity to genetic disease association, essentiality and/or dispensability. In particular our results highlight why dominant disorders are strongly associated with ohnologs, demonstrating that this has more to do with these genes tendency to be haploinsufficient rather than a result of compensation mechanisms or their so-called ‘dangerous’ nature.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-352

Retrotransposons elucidate Paraphyly within the Genus Peromyscus

Kevin Sullivan*, David Ray [1], Neal Platt [1], Robert Bradley [1 2]

[1]Biology, Texas Tech University, [2]Natural Science Research Laboratory, Lubbock, United States

Abstract: SINEs make for great phylogenetic markers due to their identity by descent, ease in determining the common ancestor, and nearly negligible risk of homoplasy. ME-Scan is a recent technique shown to elucidate relationships using 10,000's of SINEs as markers, evincing its worth as a prodigious, yet economical next-gen phylogenetic approach. Peromyscus, a speciose and ubiquitous North American genus of mice, is a well-studied mammalian taxon; yet, current phylogenies for it are mitochondrially biased and paraphyletic. We used ME-Scan on 40 taxa, each representative of an intrageneric species group, to provide the first data and character-rich phylogeny, as well as the first nuclear phylogeny, of Peromyscus.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-378

Stop and go, fast and slow: can the Covarion model of rate variation improve our understanding of phylogenies and molecular evolution?

Ashley Schoonmaker [1,*], Lauren Rodriguez, Lyndon Coghill [1], Jeremy Brown [1]

[1]Biological Sciences, Louisiana State University, Baton Rouge, United States

Abstract: The use of genome-scale data is steadily increasing in phylogenetics, giving hope to an eventual resolution of many challenging clades in the tree of life. While these data often produce results with added resolution and strong support, recent studies have shown that small methodological changes can produce strongly conflicting results. These inconsistencies can be driven by several types of analytical problems, including systematic error resulting from poor model fit. For many types of genomic data, the best-fit model appears to not be fully adequate to explain the underlying evolutionary processes. One process that is ignored by standard models, but almost certainly important to large genomic datasets, is variation in the rate of evolution at a site through time (heterotachy). The covarion model is an elegant, but largely underutilized, approach to account for heterotachy. Here we explore whether the covarion model provides a better statistical fit to phylogenomic datasets, and whether such differences have an effect on resulting inferences about phylogeny and patterns of molecular evolution. Early results suggest that the covarion is strongly preferred to standard models and that it can alter our understanding of phylogeny and molecular evolution in important ways.

Statement: All my life I have been surrounded by technology and have had contact with and have been able to explore nearly every topic in the computing industry. At the same time, I have also been fascinated in the outdoor world and how all the organisms play a part in maintaining a stable ecosystem. Because of these two interests, I have decided to study Microbiology and Computer Science at Louisiana State University. I am currently a undergraduate senior with one more year of study before completion. It is my plan to be able to join the two disciplines together by using the knowledge gained from Computer Science to analyze important biological and evolutionary genetic problems.
I have found that I enjoy solving the problems and using scripts to run statistical analysis on biological data as well as the collecting of the data in the lab. It is my goal to find a career that will allow me to unite the two tasks in order to fully analyze and understand the data of whatever project that I am working on. I am interested in learning more about computational research in the molecular genetic and evolutionary fields. I wish to attend the SMBE meeting because I think this program can assist me in broadening my understanding of the research opportunities available in molecular evolution. By participating in this program, I would have the chance to further my understanding of computational analysis of evolutionary relationships.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-431

PATTERNS OF INTROGRESSION IN BUDDING YEAST

Anne Clark [1,*], Josh Akey [2]

[1]University of Washington, Seattle, [2]Princeton University, Princeton, United States

Poster: From the large number of yeast genome sequences now becoming available, it's clear that hybridization is widespread among budding yeast from agricultural, clinical, and wild settings. However, we still understand little about the role that hybridization has played in the evolution of yeast. Identifying introgressed sequences that result from past hybridization events can help us understand the importance of hybridization in adaptation to new environments. In addition, it may allow us to find specific genes or regions of the genome in which introgression has been important. I have developed a hidden Markov model-based tool for identifying introgressed sequences in yeast. I have applied this tool to the genomes of a large number of *S. cerevisiae* strains, identifying sequences introgressed from *S. paradoxus*, S. *bayanus*, and unknown sources. I will discuss broad patterns in the amount and distribution of introgressed sequence across these strains, as well as specific examples of genes that show surprising evolutionary histories. I will also discuss preliminary results of assessing the phenotypic consequences of these introgressed sequences through experimental assays.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-428

PROTEOME OF MYXOZOAN POLAR CAPSULES: LIGHTS INTO THE EVOLUTIONARY SCENARIO OF EXTRUSION APPARATUS IN EUKARYOTES

Qingxiang Guo [1], Yang Liu [1], Xiuping Zhang [1], Bo Zhang [1], Zemao Gu [1,*]

[1]Department of Aquatic Animal Medicine, College of Fisheries, Huazhong Agricultural University, Wuhan, China

Poster:

Extrusion apparatus, which has been observed in microsporidians, dinoflagellates, ciliates, cnidarians and myxozoans, is a primary example of a complex apparatus whose origins and evolutionary history have proven difficult to reconstruct. Among the extrusion-apparatus-bearing organisms, myxozoans represent a major lineage of metazoan parasites with extremely simplified morphology and structure. Recent phylogenomic evidence suggests an evolutionary origin of myxozoan within cnidaria and for a long time the homology between cnidarian nematocysts and myxozoan polar capsules is the key to indicating the close relationship of the two lineages. However, in contrast to extensively studied nematocysts, the myxozoan polar capsules still remains poorly characterized. To gain better understanding of the structure, function of this myxozoan-specific organelle, to straightforwardly re-evaluate the relationship between myxozoans and cnidarians, and to explore the origin and evolution of extrusion apparatus, here we present the first proteome map of polar capsules from myxozoans. Major components of polar capsules isolated from three myxobolids, *Myxobolus honghuensis*, *Myxobolus wulii* and *Thelohanellus kitauei*, were characterized by using our newly developed reference species-specific proteome called Myxozoan Comprehensive Proteomic Identification Datasets (MCPID), which are derived from deep transcriptome profile and partial genome sequencing data. The MCPID facilitated the proteomic analysis by identifying 19.1%>43.8% more proteins with a maximum 84.6% of database size reduction, finally enabling the identification of 1111, 490 and 597 polar capsule proteins (PCPs) in *M. honghuensis, M. wulii and T. kitauei* respectively. Comparative proteomics show that the polar capsule proteomes are highly elaborate and include novel structural proteins and venom proteins. Extending the comparison with extrusion-apparatus-bearing organisms enables us to inspect several key questions in the assumptions about evolutionary scenario of extrusion apparatus: (1) Do the myxozoan polar capsules and cnidarian nematocysts arise de novo, or do they evolve from pre-existing structures? (2) Do they evolve from protists and predate multicellularity?

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-421

A NEW FORMULATION OF RANDOM GENETIC DRIFT AND ITS APPLICATION TO THE EVOLUTION OF CELL POPULATIONS

Yuxin Chen [1,*], Ding Tong [1 2], Chung-I Wu [1 3 4]

[1]School of Life Sciences, Sun Yat-Sen University, Guangzhou, China, [2]Biostatistics, Yale University, New Haven, [3]University of Chicago, Chicago, United States, [4]Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Poster: Random genetic drift, or stochastic change in gene frequency, is a fundamental evolutionary force that is usually defined within the ideal Wright-Fisher (WF) population. However, as the theory is increasingly applied to populations that deviate strongly from the ideal model, a paradox of random drift has emerged. When drift is defined by the WF model, it becomes stronger as the population size, *N*, decreases. However, the intensity of competition decreases when *N* decreases and, hence, drift might become weaker. To resolve the paradox, we propose that random drift be defined by the variance of "individual output", V(*k*) [*k* being the progeny number of each individual with the mean of E(*k*)], rather than by the WF sampling. If the distribution of k is known for any population, its strength of drift relative to a WF population of the same size, *N*, can be calculated. Generally, E(*k*) and V(*k*) should be density dependent but their relationships are different with or without competition, leading to opposite predictions on the efficiency of random drift as *N* changes. We apply the "individual output" model to asexual cell populations that are either unregulated (such as tumors) or negatively density-dependent (e.g. bacteria). In such populations, the efficiency of drift could be as low as <10% of that in WF populations. Interestingly, when *N* is below the carrying capacity, random drift could in fact increase as *N* increases. Growing asexual populations, especially tumors, may therefore be genetically even more heterogeneous than the high diversity estimated by some conventional models.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-419

TRACING FUNCTIONAL PROTEIN INTERACTION NETWORKS USING A 'FEATURE-AWARE' PHYLETIC PROFILING

Holger Bergmann [1], Ngoc Vinh Tran [1], Julian Dosch [1], Bardya Djahanschiri [1], Sachli Zafari [1], Ingo Ebersberger [2,*]

[1]Goethe University Frankfurt, Frankfurt, Germany, [2]Applied Bioinformatics Group, Inst. for Cell Biology and Neuroscience, Goethe University Frankfurt, Frankfurt, Germany

Poster: Tracing the phyletic distribution and, thus, the evolution of protein interaction networks across hundreds or even thousands of species calls for reliable and scalable methods for functional annotation transfer. Standard homolog or ortholog inferences resulting in so called 'phyletic profiles' do not suffice in many cases, as the functional similarity between evolutionary related sequences can decay with time. Here, we integrate the search for orthologs with a subsequent automated scoring of the pair-wise feature architecture similarity (FAS) between a protein of interest, the 'seed' and its orthologs. Features comprise, among others, functional protein domains, secondary structure elements, transmembrane domains, and low complexity regions. In detail, the feature set for each protein is stored in a directed acyclic graph (DAG) and linear source-to-sink paths are obtained via a depth-first search. In cases of overlapping, redundant features in the architecture, we determine the highest scoring linear path pair through the DAGs of the 'seed' and the orthologous protein using – where applicable – a greedy, and otherwise an exhaustive or a heuristic strategy. The resulting score of an identified ortholog serves then as a proxy for the functional equivalence to the respective 'seed protein'. We tested the feature-aware phylogenetic profiling for its performance in identifying functional equivalents to well-studied and curated proteins from model organisms to hitherto unannotated sequences in non-model organisms. This revealed specificities and sensitivities comparable to those of KAAS and BlastKoala, two state-of-the art tools for functional annotation transfer. A Shiny/R based visualization tool facilitates then an intuitive and dynamic exploration of FAS supported phyletic profiles across all proteins represented in an interaction network. As a use-case, we apply our 'feature-aware' phyletic profiling framework to investigate the evolution of 204 extracellular proteins from the human pathogen *Acinetobacter baumannii* across more than 1,000 bacterial genomes. On this basis, we identified proteins that changed their feature architecture specifically on the lineage separating *A. baumannii* from its non-pathogenic relatives. These proteins, among them a factor presumably involved in cell adhesion, serve as promising candidates for hitherto undetected virulence factors.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-418

INFERENCE OF GENE FLOW ACROSS SPATIAL LANDSCAPES

Erik Lundgren [1,*], Peter Ralph [2]

[1]University of Southern California, Los Angeles, CA, [2]University of Oregon, Eugene,OR, United States

Poster: Genetic divergence is related to geographic distance, but the exact relationship is not straightforward, and is the result of how lineages move across the landscape. Our goal is to infer these movement rates from pairwise genetic divergences at known locations. If we approximate the habitat as a graph, there are two main models to explain patterns of genetic divergence on spatial landscapes: resistance distance and coalescence time. Coalescence time is mechanistically closer to reality and allows the inference of asymmetric migration rates, but resistance distance is more computationally feasible for large graphs. Resistance distance is equivalent to coalescence time when the graph is isotropic, but in general the two measures are different. I investigate how and when the approximation of coalescent time with resistance distance breaks down. I also provide an MCMC method to infer migration and coalescence rate parameters using coalescence times.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-422

MODELING ANCESTRY-DEPENDENT PHENOTYPIC VARIANCE REDUCES BIAS AND INCREASES POWER IN GENETIC ASSOCIATION STUDIES

Shaila Musharoff [1,*], Danny Park [1], Joshua Galanter [2], Scott Huntsman [1], Celeste Eng [1], Esteban Burchard [1 3], Noah Zaitlen [1]

[1]Department of Medicine Lung Biology Center, University of California, San Francisco, San Francisco, [2]Genentech, South San Francisco, [3]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, United States

Poster: Many complex human phenotypes vary dramatically in their distributions between populations. Genetic association studies typically use estimates of ancestry, such as principal components (PCs), as fixed-effect covariates to prevent confounding caused by a dependence of phenotypic mean on ancestry. However, the current gold standard approach approach of including PC covariates in linear regression models (LR+PC) assumes that different populations have the same phenotypic variance, which may not hold for recently admixed populations. In this work we consider the possibility that populations with differences in phenotypic mean also have differences in phenotypic variance. First, we show this is the typical case under an additive genetic architecture. Then, we develop ADGLM, a likelihood-based method based on a double generalized linear model, to account for relationships between ancestry and phenotypic variance in genetic association studies. In simulations, our test ADGLM has better power than several linear regression tests that assume equal variance across groups. We observe power increases of 12 - 66% and obtain unbiased parameter estimates for data simulated with realistic effect sizes and minor allele frequency differences of 0.45. Furthermore, we show that the standard approach LR+PC can lead to inflation or deflation of p-values for tests of genetic association when population phenotypic variances differ. For example, simulated populations with minor allele frequencies of 0.05 and 0.5 produce test statistics with an inflation factor (lambdaGC value) of 1.56, which ADGLM fixes.
When applied to the Study of African Americans, Asthma, Genes and Environments (SAGE), ADGLM find significant associations of baseline lung function (FEV_1) with global ancestry proportion when either sex or BMI is a sole covariate. By contrast, LR+PC requires additional covariates (age, sex, height, and weight) for significant associations. When applied to Puerto Ricans from the Genetics of Asthma in Latino Americans (GALA II) study, ADGLM finds ancestry significantly associated with mean methylation at 8 of ~320K genome-wide probes while LR+PC finds one significant association. ADGLM also finds 44 probes significantly associated with methylation variance, which may be due to ancestry-associated environmental effects. Overall, ADGLM finds more significant associations than linear regression with PCs possibly because ancestry affects phenotypic variances as well as phenotypic means, and is promising for other genetic and gene-environment association studies.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-427

INCREASED INTEGRATION OF MTDNA PSEUDOGENES INTO THE NUCLEAR GENOME COINCIDES WITH SPECIATION OF THE HUMAN GENUS.

Konstantin Gunbin, Konstantin Popadin, Leonid Peshkin, Sofia Annis, Rebecca Ackermann, Konstantin Khrapko*

Poster: The question: human evolution - gradual process or a rapid discontinuous change? Whether human origin was a gradual process or was a result of rapid change has been a focus of intense debate. Of particular interest has been the mid-Pliocene climate change event ~2.9-2.5 Ma, which is thought to have precipitated the separation of the genus Homo as an independent lineage, i.e. separation of Humans from non-Humans (~2.8Ma). The debate mostly concerned continuity/punctuality of the fossil record, but of course the rate of the underlying genetic change is of ultimate interest/importance. Did hominid lineage experience an increased mutation rate when a large number of hominins emerged and eventually gave rise to the split between Australopitecus/Paranthropus and Homo??The obstacle: vague timing of conventional mutations. The difficulty in answering the above question lies in the way past mutations have to be timed. Conventional point mutations are assigned to specific branches of the DNA-derived phylogenetic trees. The essence of the problem is that mutations can be located within branch segments from branching point to branching point, but the exact position within the segment is principally unknown. Because the hominid DNA-derived phylogenetic tree is rather sparsely populated with branches, the precision of mutation timing is low, e.g., human-specific mutations can be positioned within ~6 My from separation from chimpanzee till branching of the Denisovans.The solution: NUMTs – mutations with an internal clock. NUMTs are insertions of mtDNA sequences into the nuclear genome. Unlike point mutation, each NUMTs actually represents a branch on the mtDNA phylogenic tree and thus its time of insertion can be determined as precise as their branching point can be positioned on the tree. In a sense, NUMTs are “mutations with an internal clock”, which is synchronized with the well-established mtDNA mutation evolution clock. By determining the NUMTs’ insertion time points, one can ask whether NUMTs were inserted uniformly over time or preferentially during certain periods of evolution, as implied by the “punctuated evolution” model.Results: Hundreds of pseudogenes have been cataloged in the human genome that have been inserted over the last ~60 My of which we considered the last 6 My. Various quality filters resulted in the selection of 18 NUMTs most suitable for phylogenetic analysis. Insertion times of these 18 NUMTs appear non-randomly distributed with one cluster positioned around 2.8Ma. While timing of insertion of individual NUMTs is imprecise, the overall probability of forming such a cluster by chance is low, which makes this observation highly statistically significant.Discussion: It is tempting to hypothesize that accelerated insertion of NUMTs is somehow linked to the speciation process. NUMTs could be either "riders", i.e., their insertion could be facilitated by the overall higher genome rearrangement activity during the speciation period, or "drivers", i.e. they may more readily get fixed in the population during speciation due to increased selective pressures. If correct, the hypothesis of accelerated pseudogenization would support the idea that evolution of our genus might have been discontinuous. Disclosure of Interest: None Declared Keywords: NUMT, human origin, phylogeny, punctuated evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-424

IDENTIFYING GENES ASSOCIATED WITH LONGEVITY IN MAMMALS BY COMPARING RATES OF EVOLUTION OF ORTHOLOGOUS GENES

Amanda Kowalczyk [1,*], Maria Chikina [1], Nathan Clark [1]

[1]University Of Pittsburgh, Pittsburgh, United States

Poster: The molecular mechanisms behind aging are of considerable interest to both the scientific community and the public at large because of the prevalence of age-related ailments such as cancer, cardiovascular disease, and type II diabetes in human populations. As human life expectancy increases, understanding the molecular origins of senescence is increasingly relevant to inform medical advances that may increase lifespan and improve quality of later life. Research has revealed some techniques to increase lifespan, such as caloric restriction and manipulation of specific genes, that provide clues to how aging occurs, but our general knowledge of genetic pathways that affect aging and age-related diseases is still incomplete. The wide range of lifespans that exist in mammal species provides an excellent dataset in which to examine the genetic mechanisms behind aging. By studying how various species have evolved different life expectancies, we can reveal the mechanisms that underlie the aging process. We used a novel method to scan the genomes of 61 mammal species to find genes whose rates of evolution are associated with longevity to identify potential aging pathways. A variety of genes were found to have evolutionary rates negatively correlated with species longevity, most notably DNA repair and cell cycle control genes, meaning that these genes are under increased evolutionary constraint in long-lived species. We also found enrichments in pathways involving cell cycle control and DNA repair in species with long lifespans. In our species of interest, body size and longevity were highly correlated. Therefore, we additionally analyzed the two phenotypes individually to identify genes associated uniquely with longevity, and such study is an interesting area for future research. Increased evolutionary constraint of cell cycle and DNA repair genes in long-lived species may indicate that control of DNA damage, both at the level of normal DNA function and at the level of cell duplication, plays an important role in defining species longevity. These results suggest that DNA maintenance capabilities may be the major determinant of lifespan.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-418

ARE SPECIATION GENES MORE LIKELY TO BE CONCORDANT WITH THE SPECIES TREE?

Richard Wang [1,*], Matthew Hahn [1]

[1]Indiana University, Bloomington, United States

Poster: Speciation genes have been found to produce hybrid incompatibilities through deleterious epistatic interactions. These interactions, dubbed Dobzhansky-Muller incompatibilities (DMIs), are thought to evolve by arising on independent lineages. While not deleterious on the genetic backgrounds on which they arose, their confluence in a hybrid generates a deleterious interaction. Because of their direct influence on speciation, it has been argued that speciation genes are more likely to have concordant gene trees; that is, their gene trees are more likely to be concordant with the species tree. While DMIs have received considerable theoretical attention, most of it has been under the assumption of a fixed tree. This assumption is ineffective when considering whether speciation genes may have evolved on discordant genealogies.

Here, we examine whether loci involved in DMIs are more likely to be concordant with the species tree by explicitly considering discordant genealogies from incomplete lineage sorting (ILS). The probability of an incompatibility arising increases with the amount of time that substitutions have to accumulate. For a fixed tree, this divergence time is assumed to be equal for each locus.But under a history of ILS, discordant gene trees have different branch lengths and topologies, giving rise to different probabilities for incompatible substitutions. Subsequently, an incompatibility has different probabilities of arising based on the gene genealogy. By considering how ILS alters the expected profile of DMIs between species we show that, contrary to expectations, speciation loci are less likely to be concordant with the species tree.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-417

METAGENOMIC ANALYSES REVEAL TWO PECTINOLYTIC ENTEROBACTERIA IN A POTATO STEM FROM A RECENT SOFT ROT OUTBREAK

Jeremy Glasner [1], Afnan Shazwan Nasaruddin [2], Brooke Babler [1], Fang Yang [1], Amy Charkowski [2], Nicole Perna [1,*]

[1]University Of Wisconsin - Madison, Madison, [2]Colorado State University, Fort Collins, United States

Poster: Two genera of the family Enterobacteriaceae, *Dickeya* and *Pectobacterium*, cause soft rot diseases in plant hosts from over 80% of known Angiosperm orders. The devastating damage these pectinolytic pathogens cause on crop plants like potato poses a serious problem for farmers worldwide. Several species of each genus lead to similar disease pathologies in the field, and accurate identification to the species and strain level is essential for outbreak surveillance and management. Until recently, *Dickeya dianthicola* was rarely reported in the United States, particularly in northern potato growing regions, where certification programs, research stations, and seed potato buyers vigorously monitor local incidence. In 2015, a *D. dianthicola* outbreak was reported in Maine. Since then, this species has been observed in multiple states. Seedling certification programs help limit spread of outbreaks using species-specific PCR-based assays to sample and quarantine infected seed lots.

We undertook the metagenomic sequencing project described here to address inconsistent results between assays using different markers for samples suspected to be linked to the *D. dianthicola* outbreak, and concerns about potential bias that might be introduced by enrichment media used to isolate bacterial strains from potato tubers and stems. Using the MiSeq platform, we sequenced both a 16S rDNA-amplicon library and a shotgun library from total genomic DNA recovered from an infected field sample. The 16S sequences were analyzed using MG-RAST. The shotgun library sequences were analyzed using Bowtie2 to map individual reads to reference genomes and de novo assemblers (Meta-Ray and Meta-spades) to generate contigs. Genes were predicted for all contigs > 500 bp using Prodigal and searched against the NCBI Microbial Genomes database using BLASTN. Assembled genomes were compared using Mauve.

Both *Dickeya* or *Pectobacterium* were identifiable at the Genus level in this field sample (72.5% and 0.65%, respectively) using the 16S data alone, and from the shotgun sequence data, we were able to identify nearly complete genomes of both. Comparative analyses using read mapping and genome alignment identified these to the strain level as *D. dianthicola* and *P. parmentieri*. The average contig coverage suggests that *D. dianthicola* (>600-fold) is considerably more abundant in this sample than *P. parmentieri* (25.5-fold). The *D. dianthicola* recovered from the metagenome assemblies is genetically distinct from the reference genomes, but it is most similar to a European strain, RNS04.9, isolated in 2004. We observed variation at one locus that explains inconsistency of one standard species-level diagnostic PCR assay. The *P. parmentieri* genome is remarkably similar to one recently isolated from a potato tuber in France. Our approach illustrates the value of using metagenomic analysis for outbreak surveillance alongside traditional diagnostic methods. It also begs the question of whether pest management strategies that result in quarantine of strains based on a small number of diagnostic markers may also result in selection for strains that evade detection while remaining effective pathogens. Further experimentation is underway to investigate whether coinfection of these strains enhances or inhibits progression of disease in susceptible host plants.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-414

SINGLE CELL TRANSCRIPTOMICS IN MICE WITH A HUMANIZED VERSION OF FOXP2

Benjamin Vernot [1,*], Gray Camp [1], Wulf Hevers [1], Barbara Treutlein [1 2], Svante Pääbo [1]

[1]Max Planck Institute For Evolutionary Anthropology, Leipzig, [2]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

Poster: Foxp2 is known to affect speech and language in humans, and is conserved in the mammalian lineage. A striking exception to this conservation are two amino acid changes which occurred and fixed in the modern human lineage, suggesting the possibility that these changes could contribute to the unique ability of humans to speak. Mouse models are consistent with this hypothesis - mice carrying the human version of Foxp2 learn faster in the presence of environmental cues, and have differences in their vocalization patterns. However, the molecular function of these two amino acid changes have been difficult to untangle, and may vary between brain regions and even cell types. To this end, we performed single cell RNA-seq on striatal and cortical neurons in humanized and wildtype mice, and identify significant expression changes associated with the humanized version of Foxp2. Differentially regulated genes are enriched for Foxp2 CHIP-seq occupancy in promoter regions and synapse-related cellular components (GO), and include genes associated with visual learning. Interestingly, many of these changes are cell-type specific, and some involve promoter-switching, suggesting interaction partners play a role in these human-specific changes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-433

PSEUDOGENES OF MTDNA SUGGEST REPEATED DISTANT INTER-SPECIES HYBRIDIZATION IN EVOLUTION OF THE HUMAN LINEAGE.

Konstantin Gunbin, Konstantin Popadin, Leonid Peshkin, Sofia Annis, Zoe Fleischmann, Natalya Markuzon, Yevgenya Kraytsberg, Rebecca Ackermann, Konstantin Khrapko*

Poster: Introduction: Increasingly, the emergence and evolution of our species is being revealed as a period characterized by genetic exchange between divergent lineages (Neanderthals, Denisovans, and other hominins that had diverged from our lineage as early as 700Ka). However, we have little knowledge on the existence or prevalence of genetic exchange during earlier (pre-1Ma) periods in human evolution as well as on the exchange with more divergent lineages. The hypothesis that the evolution of humans involved hybridization between highly diverged hominin lineages has been actively debated in recent years. Results: We present evidence supporting occurrence of hybridization within human lineage, show that it likely happened between highly divergent (~4.5My) lineages, more than one time. We use analysis of nuclear pseudogenes of mtDNA ("NUMTs"). NUMTs are considered "mtDNA fossils", as they preserve sequences of ancient mtDNA because mutational rate in the nucleus is much lower than in mtDNA. We demonstrate that a NUMT on human chromosome 5, which is shared by chimpanzee and gorilla, had descended from a mitochondrial genome that had been divergent from our ancestor's mtDNA by ~4.5% nucleotides at the time of pseudogene insertion. This implies that this pseudogene should have been created in a hominid that at that time had been diverged by about 4.5My of evolution from the hominid that at that time carried our mtDNA lineage. In order for this pseudogene and our mtDNA to end up in the same body, these two hominids should have mated (hybridized) with each other. The large divergence implies a distant interspecies (or even inter-generic) hybridization. Additionally, analysis of two other NUMTs (on Chr11 and Chr7) suggests that hybridization events occurred repeatedly. To exclude the alternative explanation by the large ancestral population size, we have shown that mtDNA divergence in contemporary ape populations does not depend on nuclear DNA-derived effective population size. Discussion: It is thought that within mammals, it takes ~2-4My to establish reproductive isolation. However, fertile inter-generic hybrids have been documented among several primates, separated for ca. 4Myr. Very recently, hybridization between Colobine genera separated by ~5.5Ma was reported to involve a NUMT scenario similar to what we had proposed human ancestors. Phylogeny consistently places the ch5 pseudogene insertion around the time of the Homo/Pan split. Intriguingly, certain hominin fossils of that epoch have been interpreted alternately as more human-like or more ape-like. Such morphological mosaicisity could potentially be explained by hybridization. Fixation of NUMTs within population should have been rather efficient, since these pseudogenes appear to have been fixed in more than one population. Thus their spread across populations might have been driven by positive selection. Indeed, NUMTs on chr5 and chr11 are located in 3' regions of functional genes. Most intriguingly, Ps11 is located 3' to the RNF141/ZNF230 gene, essential for spermatogenesis. NUMT might have served as an expression modifier for RNF141, resulting in reproductive advantage. Indeed, RNF141 demonstrates selectively driven expression shift in testis of the ancestor of hominines. Disclosure of Interest: None DeclaredKeywords: Hybridization, Hominin evolution, NUMT, mtDNA, Phylogeny

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-387

Development of Evaluation Guidelines of Test Method for IDV for Sexually Transmitted Infection

San Kim [1,*], Geun Soo Kim [1], Hye Youn Han [1], Eunji Kwon [1], Mi Suk Noh [2], Si-Hyung Yoo [1], Hyegyeong Min [1], Insu Lee [1], Chang Won Park [1], Seung-Hwa Hong [1]

[1]Medical Device Research Division, National Institute of Food and Drug Safety Evaluation, Cheongju-si, Chungcheongbuk-do, [2]Medical Supplies Evaluation Center, Korea Testing Certification, Gunpo-si, Gyeonggi-do, Korea, Republic Of

Abstract: *Chlamydia trachomatis* is usally isolated only from humans and is often infected together with other sexually transmitted diseases(STD). *C. trachomatis* is a serious disease which infects 500 million people per year globally and causes blindness in 7 to 9 million people per year due to continuous infection. Thus, it is urgent to develop (draft) guidelines on the test methods for performance evaluation of medical devices for in vitro diagnosis of bacterial STD(*C. trachomatis*) based on molecular diagnosis. In this study, in reference to the U.S. FDA Guidance, has derived essential methods for testing performance which are applicable to medical device for in vitro diagnosis of bacterial STD which are used in the field of melecular diagnosis. Because of the establish of the test method for performance evaluation of medical devices, it could provide consistent performance evaluation standards for the approval and examination of products, promote the improvement of the quality and safety of products. This will contribute to the development of the quality and safety of products. This will contribute to the development of medical device for in vitro diagnosis in Korea.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-404

THE GENOMIC ARCHITECTURE OF PREFERENCE FOR HUMAN HOSTS IN THE ZIKA AND DENGUE MOSQUITO AEDES AEGYPTI

Noah Rose*

Poster: Mosquitoes are major vectors of human diseases such as Zika, malaria, dengue fever, yellow fever, and West Nile virus. The most dangerous mosquito species are those that live in association with humans and preferentially bite humans. Interestingly, however, many vector species are ecologically variable in their associations with human hosts and habitats. This variation provides an opportunity to determine the genetic and cellular mechanisms that underlie adaptation to human hosts. *Aedes aegypti*, which transmits the Zika, chikungunya, yellow fever, and dengue viruses, comprises two major subtypes. The non-African subtype *Aedes aegypti aegypti (Aaa)* uniformly prefers human hosts, whereas the African *Aedes aegypti formosus (Aaf)* shows wide variation in attraction to humans. A previous study identified an odorant receptor associated with preference for humans in *Aaa* by comparing antennal gene expression in *Aaa* x *Aaf* F2 hybrids that differed strongly in preference. This analysis was limited to genes showing expression differences in adult antennae because the reference genome assembly available at that time was too poor for unbiased linkage mapping. The recent completion of a high quality end-to-end assembly of the *Aedes aegypti* genome allowed us to revisit this experiment and carry out an unbiased genome-wide search for host preference loci that act at any life stage or tissue (including the central nervous system). We identify clear QTL peaks that contain novel host preference candidate genes, suggesting that additional mechanisms beyond differences in odorant receptors played a role in the evolution of preference for human hosts in *Aedes aegypti*. We also identify genomic regions showing especially strong genetic differentiation between *Aaa* and *Aaf* and show that the vast majority of variants that drove the evolution of preference for human hosts in *Aaa* are likely still variable within *Aaf* today.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-403

RENT+: AN IMPROVED METHOD FOR INFERRING LOCAL GENEALOGICAL TREES FROM HAPLOTYPES WITH RECOMBINATION

Sajad Mirzaei [1], Yufeng Wu[*]

[1]University Of Connecticut, Storrs, United States

Poster: Haplotypes from one or multiple related populations share a common genealogical history. If this shared genealogy can be inferred from haplotypes, it can be very useful for many population genetics problems. However, with the presence of recombination, the genealogical history of haplotypes is complex and cannot be represented by a single genealogical tree. Therefore, inference of genealogical history with recombination is much more challenging than the case of no recombination.

In this work, we present a new approach called RENT+ for the inference of local genealogical trees from haplotypes with the presence of recombination. RENT+ builds on a previous genealogy inference approach called RENT, which infers a set of related genealogical trees at different genomic positions. RENT+ represents a significant improvement over RENT in the sense that it is more effective in extracting information contained in the haplotype data about the underlying genealogy than RENT. The key components of RENT+ are several greatly enhanced genealogy inference rules. Through simulation, we show that RENT+ is more efficient and accurate than several existing genealogy inference methods. As an application, we apply RENT+ in the inference of population demographic history from haplotypes, which outperforms several existing methods.

RENT+ is implemented in Java, and is freely available for download from: https://github.com/SajadMirzaei/RentPlus.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-401

THE MISSING LANDSCAPE OF HUMAN GENOMIC DIVERSITY IN THE ARABIAN PENINSULA

Njlaa Bakhsh [1,*], Latifa Jackson, Fatimah Jackson, christopher Cross

[1]Howard University, washington, United States

Poster: The Arabian Peninsula (AP) is the first site of human migration and habitation outside of Africa. As a major crossroad for human populations, the AP provides an opportunity to better understand early to modern changes in human demographic patterns through selections, admixture, gene flow, and migration. Dramatic climatic fluctuations have been recorded in the AP that contributed to contractions and expansions in water availability. These climatological perturbations are thought to have shaped genomic variations in this population. Recent reports indicate that a number of Arab nation-states have committed significant resources to genetically type the national population, with the overall goal of determining the degree of genomic diversity in the AP. We sought to characterize currently typed genomic variation in Arabian populations to support the rationale for our proposed analyses of Saudi Arabian genomic diversity. Interestingly, in contrast to published claims , a comprehensive search of peer-reviewed reports on genomic analysis (N=20 papers) revealed no genomic data from four national genomic projects (Qatar, Saudi Arabia, Kuwait, and The United Arab Emirates). Our analysis demonstrates that while much fanfare and presumably resources have been devoted to defining the genomic landscape of the Arabian peoples, little actual data is available to either substantiate or support such an investment.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-405

PORPOISE GENOMICS REVEALS INSIGHTS INTO MOLECULAR ADAPTATIONS TO AQUATIC TRANSITION AND THE RISE OF A NEW CETACEAN SPECIES

Xuming Zhou*

Poster: Cetaceans (whales, dolphins and porpoises) are a group of mammals fully adapted to various aquatic habitats, from oceans to freshwater rivers. We report the sequencing (106× coverage), *de novo* assembly and analysis of a finless porpoise genome. Changes in gene families and positively selected genes were identified in the cetacean genomes compared to their terrestrial relatives. By profiling bone microanatomical structure across 5 cetaceans and 23 other mammals, we identified genes with the root-to-tip substitution rate that are correlated with global bone compactness, some of which were implicated in bone mass control. We further sequenced the whole genomes of additional 48 finless porpoise individuals from seven major distributed locations and 11 sampling sites from a relatively wide geographic range across Chinese coastal waters and the Yangtze River, population genetic and species delimitation analyses suggested that Yangtze finless porpoises is a distinct species. Genes under selection between marine and freshwater porpoises (such as urea transporter 2 and angiotensin I-converting enzyme 2) are associated with renal water homeostasis, urea cycle and renin-angiotensin system, which may explain how finless porpoises are able to cope with different osmotic stress in ocean and rivers.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-409

EVOLUTION OF THE 3R-MYB GENE FAMILY IN PLANTS: A LINK BETWEEN CELL CYCLE AND ABIOTIC STRESSES

Guanqiao Feng [1,*], J. Gordon Burleigh [1 2 3], Edward L. Braun [2 3], Wenbin Mei [3 4], W. Brad Barbazuk [1 2 3]

[1]Plant Molecular and Cellular Biology Program, [2]Genetics Institute, [3]Department of Biology, University of Florida, Gainesville, [4]Department of Plant Sciences , University of California, Davis, United States

Poster: Plant 3R-MYB transcription factors are an important subgroup of the MYB super family in plants; however, their evolutionary history and functions remain poorly understood. We identified 225 3R-MYB proteins from 65 plant species, including algae and all major lineages of land plants. Two segmental duplication events preceding the common ancestor of angiosperms have given rise to three subgroups of the 3R-MYB proteins. Five conserved introns in the domain region of the 3R-MYB genes were identified, which arose through a step-wise pattern of intron gain during plant evolution. Alternative splicing (AS) analysis of selected species revealed that transcripts from more than 60% of 3R-MYB genes undergo AS. AS could regulate transcriptional activity for some of the plant 3R-MYBs by generating different regulatory motifs. The 3R-MYB genes of all subgroups appear to be enriched for Mitosis-Specific Activator (MSA) element core sequences within their upstream promoter region, which suggests a functional involvement in cell cycle. Notably, expression of 3R-MYB genes from different species exhibits differential regulation under various abiotic stresses. These data suggest that the plant 3R-MYBs function in both cell cycle regulation and abiotic stress response, which may contribute to the adaptation of plants to a sessile lifestyle.
Key words: 3R-MYB; gene family evolution; alternative splicing; intron evolution; cell cycle; abiotic stresses

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-407

THE EVOLUTION OF NA+, K+-ATPASE IN RELATION TO PREDATORY FEEDING AND LUCIBUFAGIN SEQUESTRATION IN FIREFLIES

Lu Yang [1,*], Matthew Aardema [2], Ying Zhen [3], Jamie Ding [4], Peter Andolfatto [1 5]

[1]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, [2]Sackler Institute of Comparative Genomics, American Museum of Natural History, New York, [3]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, [4]Department of Molecular Biology, [5]Lewis-Signer Institute for Integrative Genomics, Princeton University, Princeton, United States

Poster: Although conspicuous in the dark, fireflies are usually avoided by predators because they contain a group of potent bitter-tasting and toxic chemical compounds called lucibufagins. However, there are exceptions: in the genus *Photuris*, fireflies are unable to produce an effective chemical defense of their own. Instead, they have evolved the ability to capture other more toxic fireflies, prey on them and sequester the toxin for their own defense. It is unknown how the *Photuris* fireflies are not affected by lucibufagin toxicity, which inhibits Na^+, K^+-ATPase, a key enzyme used in a variety of cellular processes including neural signal transduction and muscle contraction. Here, we found the gene of Na^+, K^+-ATPase has duplicated three times in *Photuris* and the duplicates bear amino acid substitutions at key toxin-binding sites, which renders these fireflies insensitivity to lucibufagins. Furthermore, our results show a tissue-specific expression pattern of the duplicates. The most resistant copy is upregulated in the gut, the tissue exposed to the highest concentrations of lucibufagins, but not in head, where the toxin is blocked by the blood-brain barrier. This divergent expression pattern helps preserve the duplicates and lower fitness costs from negative pleiotropy. This study is an example of how next-generation sequencing can help us better understand ecology and behavior in firefly research.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-400

ANCIENT PROTEINS AND THE THRIFTY GENE HYPOTHESIS: URIC ACID'S CONTRIBUTION TO PRIMATE EVOLUTION

Jennifer E. Farrar [1,*], Lily Tran [1], Adrianna C. Brown [1], Eric A. Gaucher [1]

[1]Georgia Institute Of Technology, Atlanta, United States

Poster: The protein uricase, which breaks down the insoluble molecule uric acid, is found ubiquitously throughout each domain of life. However, certain animal lineages, including birds, reptiles, and apes, have experienced distinct evolutionary events leading to the pseudogenization of uricase. In order to understand the evolutionary process behind the inactivation of the uricase gene in apes, we have synthesized ancestral uricases by employing evolutionary models to predict the ancestral sequences of these proteins. We discovered that uricase gradually lost activity stepwise prior to its inactivation. Ancestral sequence reconstruction on two uric acid transporters, URAT1 and ABCG2, also showed that the function of the ancient transporters changed concomitantly with the ancient uricases. Further, preliminary phylogenetic analysis of the enzyme that converts xanthine to uric acid (xanthine oxidoreductase, XOR) suggests that functional constraints acting on this enzyme changed at the same time that uricase and its transporters experienced a change in function. It is unknown why these events have occurred, particularly because the buildup of uric acid in the body causes hypertension, renal disease, liver damage, and gout. However, with our collaborators, we have shown that increased uric acid levels are important for the conversion of fructose into triglycerides. These results support a recent hypothesis that frugivory is responsible for the evolution of large brain size in primates. Higher levels of uric acid would have facilitated the digestion of the fructose-rich diets of these primates, allowing more energy to be used for an increased encephalization quotient in primates.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-370

Phylogenomic analysis of sex-chromosome meiotic drive reveals divergent genomic consequences of drive in a species pair

Josephine Reinhardt [1,*], Richard Baker [2], Kimberly Paczolt [3], Gerald Wilkinson [4]

[1]Biology, State University of New York at Geneseo, Geneseo, [2]American Museum of Natural History, New York, [3]University of Maryland, College Park, United States, [4]Biology, University of Maryland, College Park, United States

Abstract: Meiotic drive alleles violate the law of equal segregation and as a result, have an advantage over other alleles at the same locus. In the case of X-linked drive alleles, the sex-ratios produced by male carriers are distorted in favor of females, as Y-bearing sperm fail to develop. In theory, drive loci can be stably maintained as single loci, but they are often found as part of large - even chromosome wide - inversion polymorphisms. Here, we report an analysis of transcriptomes collected from a family of flies (Diptera: Diopsidae) known to carry meiotic drive polymorphisms in multiple species. Comparing sequences of hundreds of expressed genes from the testes of males with distorted and standard sex-ratios in two species, we found evidence of both situations. In one species (*Teleopsis dalmanni*), meiotic drive was part of a chromosome-wide haplotype shared by individuals from multiple collections. This drive system was associated with widespread divergence in gene sequence and expression. In a closely related sympatric species (*T. whitei*), meiotic drive had little detectable impact on gene expression or sequence divergence, despite the fact that the sex-ratio phenotype was detected at similar, apparently stable frequencies in the two species. The evidence is consistent with a model wherein X-linked meiotic drive exists as a freely recombining polymorphic locus in *T. whitei* but as part of an ancient chromosome-wide inversion in *T. dalmanni.*

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS16

The fitness landscapes of a yeast tRNA gene in multiple environments: G×E is pervasive yet simple

Jianzhi Zhang [1,*]

[1]University of Michigan, Ann Arbor, United States

Abstract: Fitness landscapes describe the genotype-fitness relationship and represent major determinants of evolutionary trajectories. However, the multiple genotypic permutations, coupled with the difficulty of measuring fitness, have hindered the empirical determination of fitness landscapes. Combining precise gene replacement and next-generation sequencing, we quantified the Darwinian fitness at 30°C in a rich medium for over 26,000 yeast strains, each carrying a unique variant of a single-copy Arginine tRNA gene at its native genomic location. Approximately 2% of all 207 possible single point mutations show over 5% fitness advantages, while 31% show over 5% deleterious effects. Approximately 57% of mutation pairs exhibit significant epistasis, with a strong negative bias except when the mutations occur at Watson-Crick paired sites in the tRNA secondary structure. Fitness is broadly correlated with the predicted fraction of correctly folded tRNA molecules, revealing a biophysical basis of the fitness landscape. To study how the fitness landscape of the tRNA gene varies among environments, we measured the landscape in three additional environments. We found that the same mutation almost always has different fitness effects in different environments, indicating pervasive genotype by environment interactions (G×E). Nevertheless, the observed G×E follows a simple rule that the fitness effect of a (deleterious) mutation in an environment is proportional of the fitness effect of deleting the tRNA gene (i.e., gene importance) in the environment. This rule allows predicting the fitness landscape of the tRNA gene in any environment as long as it has been measured in one environment and the relative gene importance in the two environments compared is known. Taken together, our high-throughput mapping reveals relatively simple rules underlying the seemingly complex tRNA fitness landscapes, giving hopes for understanding and predicting fitness landscapes of other genes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-362

Replay Experiments Uncover the Effect of Historical Contingency on Evolutionary Outcomes

Ryan Vignogna [1,*], Sean Buskirk [1], Gregory Lang [1]

[1]Biological Sciences, Lehigh University, Bethlehem, United States

Abstract: Epistasis plays a critical role in constraining evolutionary outcomes. However, we are limited in the few studies of epistasis that exist, and fewer still where epistasis directly relates to dynamics of adaptation. We have identified synergistic epistasis in a population for which we have high-resolution knowledge of the dynamics of genome sequence evolution. In this experimentally-evolved yeast population, a high-fitness *ste12* lineage was abruptly outcompeted by a lineage containing 11 mutations. The earliest of these mutations, *iqg1*, existed at low frequency for hundreds of generations before emerging with the cohort at generation 500. We hypothesize that the fate of the *iqg1* lineage depended on the accumulation of early, non-adaptive mutations and thus is an example of historical contingency. We performed evolutionary "replay experiments". Population samples from six time-points, during which the *iqg1* lineage persisted at low frequency, were each used to seed 48 replicate populations. Through a fluorescent reporter we tracked the frequency of the *ste12* lineage until it fixed or went extinct. We find the likelihood of the *iqg1* lineage outcompeting the *ste12* mutation increases through the six time-points, indicating contingency playing an important role in the lineage's success. We are currently sequencing to determine the spectrum of beneficial mutations in our replay experiments. Our results show that through epistatic interactions, a low-frequency lineage may be able to follow an evolutionary path to higher fitness that is not easily accessible to the majority lineage. This provides new insight into the balance between chance and contingency in determining evolutionary outcomes.

Expanded summary*: We lack understanding in how epistatic interactions arise and how epistasis affects dynamics of adaptation. This lack of knowledge hinders our ability to interpret mutation patterns in evolved genomes and remains a blockade in predicting evolutionary processes. Our replay experiments sought to determine how epistasis affects evolutionary outcomes. Given the dynamics of the *iqg1* lineage, where the *iqg1* mutation stayed at low frequency for hundreds of generations before finally outcompeting the high-fitness *ste12* mutation with a group, or cohort, of 10 other mutations, we hypothesized that the first mutations to occur in this cohort increased the fitness advantage of subsequent mutations. Results so far show that the likelihood of the *iqg1* lineage winning increases over the six replay time-points. Next-generation sequencing is currently being done to determine the range of beneficial mutations in these lineages. Fitness data has been done on the original population and similar work will be done for these sequenced lineages as we continue to determine the role of epistasis in the success *iqg1* lineages. Despite the focus on a single lineage our work broadly addresses the following: given a dominant lineage containing a beneficial mutation and a less-fit minority with higher evolutionary potential, which will ultimately succeed and why? This is a vast departure from other studies of epistasis because they typically ignore the dynamics of adaptation. Our results provide unique detail into the patterns of fitness effects and epistatic interactions in adaptive evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-369

Mycorrhizal fungi associated with marri trees in Western Australia: a comparison across a disturbance gradient

Sarah Sapsford*, Trudy Paap, Giles Hardy [1], Treena Burgess [1] and Murdoch University

[1]Murdoch University, Perth, Australia

Abstract: Marri (*Corymbia calophylla*) is a keystone tree in Western Australia, however, a canker disease caused by the fungus *Quambalaria coyrecup* is devastating marri throughout much of its native range. Disease incidence is high in areas of anthropogenic disturbance such as roadside reserves, especially those adjacent to cleared land. The pattern of decline and symptoms suggest a complex interaction of factors including nutrient imbalance and loss of mycorrhizal fungi. These fungi are beneficial in that they are critical for nutrient acquisition and maintain the health of trees. The aim of this study was to determine changes in communities of mycorrhizal fungi in marri along an anthropogenically disturbed gradient. Seventeen sites were surveyed. Each site consisted of a disturbance gradient of 3 transects: remnant stand of marri bordering cleared land and a road, a forest edge, and the middle of an intact forest. Soil was collected from ten trees along each transect and the soil was used in a glasshouse bioassay trial where seedlings were grown for 6 months. The marri seedlings were harvested and the fine roots collected. High-throughput sequencing was carried out on the fine roots from the glasshouse trial. Mycorrhizal communities changed across the gradient from the seedling baits. These changes where both in taxonomic groups and functional composition and demonstrated the community assemblage from the disturbed edge was unique from the other transects. It is possible that communities are changing due to edaphic selection pressures and anthropogenic pressures present along the disturbed edge.

Expanded summary*: The decline of *Corymbia calophylla* (marri) has been increasing over recent years. Marri is an important tree species in Western Australia (WA); they are an important food source and nesting location for the endangered Carnaby's black cockatoo (*Calyptorhynchus latirostris*) (Taylor et al. 2012). Many other bird species and insects are also attracted to its nectar and pollen; when marri are flowering heavily, birds located in areas where fruit is cultivated are much less likely to feed on farmed fruit (Cunningham 1998). Marri is also one of the major honey plants in southwest WA and beekeepers are heavily dependent on it (Powell 1990). In addition, marri is used in the furniture industry due to its unique gumlines it produces in furniture (Forest Industry Statement 2004). Therefore, the decline of this species has devastating effects on wildlife habitats as well as the economy.

A canker caused by the fungus *Quambalaria coyrecup*, believed to be endemic to WA, is devastating many marri stands in the southwest of WA (Paap et al. 2008). As this fungus has only been described in the last decade, there is much we still don't understand about the disease it causes. In addition, it has been observed that incidences of *Q. coyrecup* are much greater in areas affected by anthropogenic activities such as parks, small remnants, peri-urban areas, cleared land used for farming and/or grazing and roadsides (Paap et al. 2016). It is possible that a combination of factors is predisposing marri to canker disease. For example, remnant stands of marri trees on road edges that border cleared land (usually a farm) have extremely high incidences of canker. However, across the road (on the opposite road edge) where there is a forest block, the incidence of canker dramatically decreases and reaches 0% incidence on trees that are at least 20 m away from the road edge (S. Sapsford, pers. obs.). Predisposing factors leading to the decline of marri may include the use of pesticides and herbicides on these road edges (either by local council spraying road edges or run-off from adjacent farmland), soil pathogens such as *Phytophthora* spp., and changes in climate. These factors can have detrimental effects on concentrations of nutrients in the soil, soil composition, soil pH and, most importantly, microbial communities in the soil.

The progression of the decline strongly suggests a breakdown in the ability of the trees to maintain nutrient balance and we strongly believe that mycorrhizal fungi play a role in this process. Previous preliminary studies with *Eucalyptus gomphocephela* have shown marked differences in the mycorrhizal communities between healthy and declining trees (Scott et al. 2013), however, no studies have been undertaken across a decline gradient. Mycorrhizal fungi are known to be more important for the uptake of vital nutrients on low fertility sites and are essential for below-ground carbon storage. Thus, any change in the mycorrhizal community would result in critical changes in nutrient uptake and tree health as evidenced by declining crown vigour. It is possible that the use of pesticides and herbicides are causing a decline in the diversity of mycorrhizal fungi in disturbed sites, and thus predisposing marri to canker. It is likely that mycorrhizal fungi are a critical component in marri decline in the southwest of Western Australia.

Aims

The main objective of this project is to tease out the different stresses that are predisposing marri to canker disease with a particular focus on below ground microbial communities.

Significance

Because very little is known about the communities of mycorrhizal fungi associated with marri, this project will further our understanding of Australasian fungi and promote its conservation. However, sequence platforms used for taxonomic identification are heavily weighted with northern hemisphere species making identification for southern hemisphere species difficult (Bik et al. 2012; Lindahl et al. 2013; Nilsson et al. 2011), and this project will add substantially to this hole in the database as I will build a library of fungal species found in WA. This project will provide critical baseline data about the mycorrhizal fungi associated with marri and their ecological importance i.e. whether changes in these associations are a predisposing factor to canker disease. This project will provide critical new information about the role that mycorrhizal communities play in tree declines to researchers and others in the mycological community.

References

Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK (2012) Sequencing our way towards understanding global eukaryotic biodiversity. Trends Ecol Evol 27:233-243. doi:http://dx.doi.org/10.1016/j.tree.2011.11.010

Cunningham I (1998) The trees that were nature's gifts. The Environmental Printing Company, Maylands, Western Australia

Forest Industry Statement (2004) WA timber industry 2004-2013. Government of Western Australia,

Lindahl BD, Nilsson RH, Tedersoo L, Abarenkov K, Carlsen T, Kjøller R, Kõljalg U, Pennanen T, Rosendahl S, Stenlid J, Kauserud H (2013) Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide. New Phytol 199:288-299. doi:10.1111/nph.12243

Nilsson RH, Tedersoo L, Lindahl BD, Kjøller R, Carlsen T, Quince C, Abarenkov K, Pennanen T, Stenlid J, Bruns T, Larsson K-H, Kõljalg U, Kauserud H (2011) Towards standardization of the description and publication of next-generation sequencing datasets of fungal communities. New Phytol 191:314-318. doi:10.1111/j.1469-8137.2011.03755.x

Paap T, Burgess TI, Calver M, McComb JA, Shearer BL, Hardy GES (2016) A thirteen-year study on the impact of a severe canker disease of *Corymbia calophylla*, a keystone tree in Mediterranean-type forests. For Pathol:1-7. doi:10.1111/efp.12292

Paap T, Burgess TI, McComb JA, Shearer BL, St J. Hardy GE (2008) *Quambalaria* species, including *Q. coyrecup* sp. nov., implicated in canker and shoot blight diseases causing decline of *Corymbia* species in the southwest of Western Australia. Mycol Res 112:57-69. doi:http://dx.doi.org/10.1016/j.mycres.2007.10.005

Powell R (1990) Leaf and branch: trees and small shrubs of Perth. Department of Conservation and Land Management, Western Australia

Scott PM, Shearer BL, Barber PA, Hardy GESJ (2013) Relationships between the crown health, fine root and ectomycorrhizae density of declining *Eucalyptus gomphocephala*. Australasian Plant Pathology 42:121-131. doi:10.1007/s13313-012-0152-4

Taylor K, Andjic V, Barber P, Hardy G, Burgess T (2012) New species of *Teratosphaeria* associated with leaf diseases on *Corymbia calophylla* (Marri). Mycological Progress 11:159-169. doi:10.1007/s11557-011-0738-1

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-387

Theory of measurement for site-specific evolutionary rates in amino-acid sequences

Dariya Sydykova [1,*], C O. Wilke

[1]University of Texas at Austin, Austin, United States

Abstract: Theory of measurement for site-specific evolutionary rates in amino-acid sequences

Dariya K. Sydykova and Claus O. Wilke

Many applications require the calculation of site-specific evolutionary rates from alignments of amino-acid sequences. For example, catalytic residues in enzymes and interface regions in protein complexes can be inferred from observed relative rates. While numerous approaches exist to calculate amino-acid rates, however, it is not entirely clear what physical quantities the inferred rates represent and how these rates relate to the underlying fitness landscape of the evolving protein. Further, amino-acid rates can be calculated in the context of different amino-acid exchangeability matrices, such as JTT, LG, or WAG, and again it is not known how the choice of the matrix influences the physical interpretation of the inferred rates. Here, we develop a theory of measurement for site-specific evolutionary rates, but analytically solving the maximum-likelihood equations for rate inference performed on sequences evolved under a mutation–selection model. We demonstrate that the measurement process can only recover the true expected rates of the mutation–selection model if rates are measured relative to a naïve exchangeability matrix, in which all exchangeabilities are equal to one. Rate measurements using other matrices are quantitatively close but not mathematically correct. Our results demonstrate that insights obtained from phylogenetic-tree inference do not necessarily apply to rate inference, and best practices for the former may be deleterious for the latter.

Expanded summary*: Different sites in a protein evolve at different rates [1,2]. The heterogeneity in rates within a protein sequences is caused by the interplay of functional and structural constraints [3]. For instance, active sites are generally very conserved [4,5]. The protein core tends to be more conserved than the surface, presumably because mutations in the core are more likely to disturb the protein structure [6,7]. Because the evolutionary rates correspond to structurally and functionally important sites, having a reliable and accurate method for inferring site-wise rate of evolution is essential. Specifically, in most viral populations, proteins evolve very rapidly. In influenza, mutations in one site of a surface protein hemagglutinin allow the virus to escape host antibodies and propagate. Thus, detecting site-wise rates of evolution can be crucial to the efforts of viral surveillance and control.

Many methods to infer site-wise rate have been developed over the years. These methods employ a substitution matrix, which captures exchangeabilities between all pairs of amino acids. The substitution matrices are made by analyzing large protein data sets and even protein sequences specific to an organism or an organelle. However, it remains an open question which substitution matrix is the most suitable for site-wise rate inference. When we measure site-wise rate, rate is defined as a scalar in front of the substitution matrix. The substitution matrix serves as a ruler by which we measure the rate of evolution at a site. Depending on what ruler or substitution matrix we use, the inferred rate changes. We developed a theory that shows the effect of the substitution matrix on the inferred site-wise rate. We demonstrate that only a naïve exchangeability matrix, in which all exchangeabilities are equal to one, can recover the true expected rates. We also demonstrate that inference with the true substitution matrix yields site-wise rate of 1. Finally, we demonstrate that the current best-practice matrices (JTT, WAG, and LG) in phylogenetic inference do not recover correct site-wise rates. Along with the mentioned results, our analytical derivations allow for further insights into models of molecular evolution.

There are two ways to measure the rate of evolution in a protein coding sequence. One by calculating the rate of evolution in codon sequences, and the other by calculating the rate of evolution in amino acid sequences. We recently demonstrated that the two inference methods produce comparable rates [8]; however, our current analytical work can establish a direct mathematical link between the two frameworks. We can directly address the issue with inferring rates with amino acid models from codon sequences. Finally, our calculations allow us to incorporate mutation rates into the inference of site-wise rate. We can test different assumptions about the mutation rates and their effect on the rate of evolution at a site.

References

1. Kimura, M. and Ohta, T. (1974). On some principles governing molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.*, 71:2848-2852

2. Perutz, M. F. et al. (1965). Structure and function of haemoglobin: II. Some relations between polypeptide chain configuration and amino acid sequence. *J. Mol. Biol.*, 13:669-678.

3. Echave, J. et al. (2016). Causes of evolutionary rate variation among protein sites. *Nature Rev. Genet.*, 17:109-121.

4. Jack, B. R. et al (2016). Functional sites induce long-range evolutionary constraints in enzymes. *PLOS Biol.* 14:e1002452.

5. Dean, A. M. et al (2002). The pattern of amino acid replacements in alpha/beta-barrels. *Mol. Biol. Evol.* 19:1846-1864.

6. Franzosa, E. A. and Xia, Y. (2009). Structural determinants of protein evolution are context-sensitive at residue level. *Mol. Biol. Evol.* 26:2387-2395.

7. A. Shahmoradi, D. K. Sydykova, S. J. Spielman, E. L. Jackson, E. T. Dawson, A. G. Meyer, C. O. Wilke (2014). Predicting evolutionary site variability from structure in viral proteins: buriedness, packing, flexibility, and design. *J. Mol. Evol.* 79:130–142.

8. D. K. Sydykova, C. O. Wilke (preprint). Calculating site-specific evolutionary rates at the amino-acid or codon level yields similar rate estimates. P*eerJ Preprints* 5:e2739v1.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-377

Whole-Genome Analysis of Polymorphism in a Bdelloid Rotifer Adineta vaga

Olga Vakhrusheva [1,*], Yan Galimov [2], Elena A. Mnatsakanova [3], Evgeny Gerasimov [1 3], Tatiana V. Neretina [4], Aleksey A. Penin [1 5], Maria D. Logacheva [1 5], Georgii A. Bazykin [1 5 6], Alexey S. Kondrashov [5 7]

[1]Institute for Information Transmission Problems, [2]Koltsov Institute of Developmental Biology, [3]Biological faculty, [4]Pertsov White Sea Biological Station, [5]Belozersky Institute for Physico-Chemical Biology, M.V. Lomonosov Moscow State University, Moscow, [6]Skolkovo Institute of Science and Technology, Skolkovo, Russian Federation, [7]Department of Ecology and Evolution, University of Michigan, Ann Arbor, United States

Abstract:

Transition to asexual reproduction is often regarded as an evolutionary dead-end. However the mere existence of ancient asexuals challenges this point of view. Bdelloid rotifers are a large group of putative ancient asexuals that have presumably abandoned sexual reproduction tens of millions of years ago. Neither males nor meiosis have ever been observed in this clade. In line with that genome structure of bdelloid rotifers seems to be incompatible with conventional meiosis. However several studies based on the sequences of individual genes report that genetic exchange appears to occur in bdelloid rotifers. To address the possibility of recombination in a bdelloid rotifer Adineta vaga on a whole-genome scale we sequenced genomes of 10 A. vaga clonal lineages.

Genomes of bdelloid rotifers pose substantial challenges for standard genome assembly approaches due to high degree of heterozygosity and atypical genomic structure devoid of homologous chromosome pairs. We obtained the draft genome assembly for one of 10 clonal A.vaga lineages using a combination of short insert paired-end libraries sequenced on the MiSeq platform and large insert mate-pair libraries sequenced on the HiSeq platform. The resulting assembly is used as a reference for downstream analysis of other sequenced samples.

We employ reference-based mapping approaches to look for recombination signatures in the whole genome polymorphism data for 10 A. vaga individuals. A decay of LD with physical distance is observed in A. vaga population suggesting that A. vaga might engage in some form of sexual reproduction.

This work was supported by RFBR grant No. 16-34-01303.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-379

Tetrahymena thermophila shows increased evolvability following sexual reproduction

Jason Tarkington* and Zufall Lab

Abstract: Understanding the mechanisms that generate genetic variation, and thus contribute to the process of adaptation, is a major goal of evolutionary biology. Tetrahymena thermophila is a ciliate with an unusual genetic feature, called phenotypic assortment, which may allow for an increase in the amount of genetic variation following sex, thereby increasing its evolvability. To test this hypothesis, I compared the rate of adaptation in T. thermophila populations that were allowed to undergo phenotypic assortment to those that were not. These populations were maintained at two different temperatures and fitness was measured every 25-50 generations for approximately 1000 generations. Under some environmental conditions, the populations that underwent phenotypic assortment adapted more quickly than those that did not. This suggests that the additional genetic variation generated by phenotypic assortment can increase the rate of adaptation under certain conditions.

Expanded summary*: *Tetrahymena thermophila* is a free-living, single-celled, facultatively sexual ciliate with two nuclei: a silent germline micronucleus nucleus (MIC) and a transcriptionally active somatic nucleus (MAC) (Merriam and Bruns 1988). The MAC gets destroyed after sex and a new one gets created from the mitotic product of the new zygotic MIC (figure 2). The MAC contains ~45 copies of every chromosome (~225) and divides amitotically (Orias and Flacks 1975, Eisen et. al. 2006). This means that the chromosomes do not line up and segregate as they do during normal metaphase and anaphase. Instead the content of the genome is divided randomly between the daughter cells (Karrer 2012). Over time this will result in one or the other parental allele being lost entirely until the entire genome, except for *de novo* mutations, is homozygous for one or the other parental allele (Sonneborn 1974). This process is known as phenotypic assortment and is contrasted with normal mitosis in figure 3. Amitotic division should result in increased genetic variance following sex and we know that the rate at which a population adapts is equal to its genetic variance in fitness (Fisher 1930).

Models have shown the novel genetic architecture of ciliates results in population genetics that differ from canonical population models (Morgens et al. 2014). Other models have shown the genetic architecture and phenotypic assortment in *Tetrahymena thermophila* allows asexual lineages to slow Mueller's ratchet (J. West pers. comm.). Additionally several studies have claimed the genetic architecture of ciliates drives rapid gene and protein evolution (Zufall et al. 2006, Gao et al. 2014). My central aim is to experimentally test the hypothesis that *T. thermophila* genome architecture and phenotypic assortment affects the dynamics of adaptation and the consequences of sex. I am testing the hypothesis that following sex phenotypic assortment increases the adaptive response in the MAC. In addition I will address issues of macronuclear inheritance and adaptive constraints by assessing how much of the evolved phenotype is lost following sex and by comparing the fitness of evolved lineages to their progeny under varying environmental conditions. This will contribute greatly to our understanding of when and how sex is advantageous in *T. thermophila.*

Phenotypic assortment is likely to increase the genetic variance during the vegetative growth period following sex. The DNA content of an asexually dividing organism is not normally considered to be plastic, changing only through various mutational processes, but in *T. thermophila* a single MAC genotype can give rise to a huge number of alternative genotypes, a property that I will refer to as genotypic plasticity. In accordance with Fisher's fundamental theorem I predict that this increased genetic variance will increase the rate of adaptation of *T. thermophila* following sex. The high chromosome number in addition to recombination between homologous chromosomes creates minimal physical linkage between loci allowing for differential assortment of each allele. This results in a large number of possible combinations of parental alleles being produced in the vegetative progeny from a single mating. With selection acting on this variation, alleles and combinations of alleles will come to dominate in environments where they are advantageous, increasing the fitness of the population. We can determine whether phenotypic assortment will indeed have this effect by comparing

the slopes of the fitness trajectories of progeny and parental populations. If phenotypic assortment increases evolvability, then the rate of fitness increase in a population derived from a single sexual progeny will be greater than that derived from an asexual individual.

Sex, despite its costs, is ubiquitous in nature and explaining this is a major challenge for evolutionary biologists (Kondrashov 1993). One of the oldest and most robust theories is that it provides an indirect benefit by increasing variation in the population thereby allowing selection to operate more effectively to increase the population fitness (Weismann 1889). This hypothesis can be contrasted with the direct benefits hypothesis in which sex increases the fitness of the parent or progeny directly (Kondrashov 1993). Indirect benefits have been demonstrated in several systems. For example, researchers using *Chlamydomonas* have provided evidence that sex increases the rate of adaptation of the population by increasing genetic variation among offspring (Colegrave et al. 2001). However, there is no reason to expect a single recombinant progeny would be any fitter than their parents. In contrast, we hypothesize that in *Tetrahymena thermophila* a single sexually produced progeny will demonstrate greater evolvability than either parent.

This would be a particularly interesting benefit of sex because it would be an indirect benefit as it takes many asexual generations and the action of selection for the benefit to manifest. However, because it is acting on the vegetative growth of a single cell and not an entire sexual population it could also be thought of as a direct benefit. In many ways one can treat all vegetative growth following sex as a single individual because they all share the same germline genome. From this perspective the direct benefit of sex is the resetting of the MAC and the increased evolvability and eventual fitness benefits that this provides to the vegetative growth of an individual cell.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-372

Community Detection over Identity-by-Descent Network Reveals Evidence of Multiple Founder Effects in a Patient Population from New York City

Gillian Belbin [1,*], Benjamin Glicksberg [2], Noam Beckmann [2], Genevieve Wojcik [3], Danny Park [4], Elena Sorokin [3], Ariella Cohain [2], Muh-Ching Yee [5], Steve Ellis [1], Adam Auton [6], Judy Cho [1], Ruth Loos [1], Girish Nadkarni [1], Noah Zaitlen [7], Chris Gignoux [3], Eimear Kenny [1]

[1]The Charles Bronfman Institute of Personalized Medicine, [2]Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, [3]Department of Genetics, Stanford University, Stanford, [4]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, [5]Department of Plant Biology, Carnegie Institution for Science, Stanford, [6]Department of Genetics, Albert Einstein College of Medicine, New York, [7]Department of Medicine, University of California, San Francisco, San Francisco, United States

Abstract: Shifting patterns of demography can profoundly impact the genetic landscape of human populations. In particular, founder effects can cause narrowing of the allele frequency spectra, and can drive potentially deleterious variants to appreciable frequencies. Thus, understanding signatures of founding is important for biomedical research, and for population screening to improved healthcare management.

One signature of founding is the presence of elevated levels of genetic sharing of haplotypes inherited Identical-by-Descent (IBD) from common ancestors. We use IBD-sharing, inferred from genotype data in >24,000 participants from a diverse biobank in New York City (Bio*Me*), to re-construct a large-scale pedigree of distant genealogical relationships in the form of a network. Using a machine learning method for community detection we find that 83% of participants fall into one of 36 communities; reflecting recent, shared ancestry at self-reported country and regional levels.

We observe evidence of founder effects in the form of elevated IBD in multiple communities including the Ashkenazi Jewish (mean pairwise sum of sharing: 28.2cM; population level probability of sharing 85%), and Finnish (11.6cM; 75%) communities; both of which are known founder populations. Others that exhibit high levels of IBD include the Puerto Rican (21.9cM; 82%), Dominican (14.7cM; 49%), Garifuna (120.5cM; 97%), Colombian (20.2cM; 75%), Ecuadorian (16.8cM; 49%) and Kenyan (11.4cM; 75%); each showing elevated sharing compared to well-studied, out-bred populations like non-Jewish European Americans (4.2cM; 3%) and African Americans (5.7cM; 4%).

Characterization of population structure using genomics can reveal signatures of founding that may highlight important populations to target for genomic medicine.

Expanded summary*: In population genetics, the ‘founder effect’ is the loss of genetic variation that occurs when a new population is established by a relatively small number of individuals. This process results in pronounced genetic drift of variants relative to the parent population. In this scenario, potentially deleterious alleles can segregate at appreciable frequencies and these alleles can contribute substantially to population-specific disease burden. Consequently, in genetic research much attention has been paid to populations that, due to geographical or cultural isolation, or migration and diaspora, share signatures of a founder effect. In concert with the majority of genomic research, much of the focus has been on founder populations of European descent. Founder populations like Iceland, Sardinia, Finland, and U.S. population isolates like the Amish and Hutterites have been extraordinarily important for advancing Mendelian and complex trait genetics.

Little is known, however, about the degree of founding in many non-European populations; meaning that many populations that have the potential to benefit greatly from genomic medicine remain understudied. Much like in Europeans, a better understanding of founder effects in populations with heritage from Africa, Asia, the Americas and Oceania may serve as bridging models to enable genomic discovery in non-European populations. However, identifying these founder effects can be challenging due to the relative underrepresentation of these population in genomic research databases, and that the underlying processes of gene flow do not necessarily correlate perfectly with the cultural, ethnic and regional labels that are typically used in research to categorize populations into discrete groups.

To address this problem, we leverage a genomic database of highly globally diverse participants from a large urban biobank in New York City. Using self-reported race/ethnicity, information about country-of-origin, and population genetic analysis, we examine for signatures of founder effects. Our method groups individuals based on statistical enrichment of Identity-by-Descent (IBD) sharing. IBD-haplotype sharing is present only between individuals with recent, shared genealogy. This also means that IBD-sharing is present at elevated levels in populations that have undergone recent bottlenecks. Using IBD allows us to both group individuals based on their genetic relatedness and to detect signatures of founder effects within these groups.

By applying our method to a highly diverse biobank from New York City, with representation from over 133 countries in the world, we are able to detect the presence of multiple communities of individuals of recent shared ancestry. Membership of these communities correlates strongly with various self-reported labels of ethnicity and region of origin. Some of these communities are highly enriched for IBD sharing, suggesting that they represent founder populations. The majority of these communities are of non-European descent and remain broadly understudied within the genomics community.

Because founder populations have the potential to harbor common deleterious alleles, identifying these populations through our method allows us then to target them to discover variants contributing to disease. For example we identified a variant in the Puerto Rican ancestry community which underlies a disorder characterized by musculoskeletal disease. While rare globally, this variant is at 2% within the Puerto Rican community and thus contributes substantially to the musculoskeletal disease burden within this group.

Our research therefore provides a framework for founder population research in the absence of accurate population labels, and helps to further genomic medicine efforts in non-European populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-381

Maps of effective migration as a summary of human genetic diversity.

Benjamin Peter [1,*], Desislava Petkova [1], Matthew Stephens [1], John Novembre [1]

[1]University of Chicago, Chicago, United States

Abstract: A dominant pattern of genetic diversity in humans is that geographically proximal populations are generally more genetically similar to one another; however, departures from the basic relationship of genetic similarity and geographic distance can arise due to persistent geographical features such as such as mountains, oceans, or deserts or other factors that elevate/depress gene flow across space. To provide more insight into how genetic differentiation is distributed geographically in humans, we examine the fine-scale genetic structure of humans. We use a non-homogeneous isolation-by-distance model to produce maps that represent the spatial structure of human genetic diversity using a recently developed method (EEMS, Estimation of Effective Migration Surfaces). We apply EEMS on global, continental, and sub-continental scales, analyzing genetic data from 8,740 individuals from 469 geographically localized populations, obtained from 24 different source studies. We find that fine-scale geographic features moderate gene flow and lead to non-homogeneous isolation by distance as a general feature of human genetic diversity. We also find the model breaks down in locations where social structure or the presence of geographically overlapping ethnic groups, significantly contribute to genetic structure. Overall, our results suggest that diversity patterns are consistent and primarily shaped by the signature of the Out-of-Africa expansion, but that migration rates are strongly influenced by geography and local events.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-380

A genome-wide test for hybrid speciation in macaques

Laurie Stevison [1,*], Ben Evans [2], Stephen Sefick [1], Don Melnick [3], Jeff Wall [4]

[1]Biological Sciences, Auburn University, Auburn, United States, [2]Biology, McMaster University, McMaster, Canada, [3]Ecology, Evolution, and Environmental Biology, Columbia University, Columbia, [4]Institute for Human Genetics, University California San Francisco, San Francisco, United States

Abstract: Studies based on phenotypic variation estimated that >10% of primate species naturally hybridize; however, the past 50 years have shown us that genetic analysis often uncovers contradictory results with undetected hybridization. Additionally, Schumer et al (2014) found that only four taxa pass their three defined criteria for evidence of hybrid speciation despite the many proposed cases found in the literature. Here, we have sequenced two species of endangered macaques to address the putative hybrid origin of *Macaca arctoides*. This species has unique phenotypic and genetic features, including species-specific genital morphology, that make it likely to have evolved under strong selection for reproductive isolation. Based on studies reporting incongruence between mitochondrial and autosomal genealogies, *M. arctoides* was proposed to have evolved via an ancient hybridization event between the ancestor of the modern Fascicularis species group and an ancestor of the Sinica species group. We test the putative hybrid origin hypothesis using genomic data from five species of macaques from both species groups, including two we sequence here for the first time, *M. arctoides* (21.2x) and *M. assamensis* (13.8x). For these two species, we identified 5.6 and 5.9 million SNPs, respectively, as compared to the rhesus reference genome. Using these high quality SNP variants, we conducted a sliding window analysis for hybridization via the four-species (aka ABBA-BABA) test using baboons as an outgroup. We analyzed evidence of hybridization independently for autosomes, the X-chromosome and the mitochondrial genome. Interestingly, we find more evidence of hybridization on the X-chromosome than the other genomic regions. Finally, recent studies in yeast have shown that hybrid speciation leads to decreases in genomic stability. To explore this possibility in primates, we characterized structural variation in *M. arctoides* and the parental lineages. We further use these SV data to determine if there have been any large-scale differences in karyotype in the evolution of these species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-374

Anchored phylogenomics resolves deep phylogeny of cichlid fishes and reveals ancestral inter-tribal hybridisation

Iker Irisarri [1], Pooja Singh [2,*], Stephan Koblmüller [2], Paolo Franchini [1], Julián Torres-Dowdall [1], Frederico Henning [1], Christoph Fischer [3], Alan R. Lemmon [4], Emily Moriarty Lemmon [5], Gerhard G. Thallinger [3], Christian Sturmbauer [2], Axel Meyer [1]

[1]Biology, University of Konstanz, Konstanz, Germany, [2]Zoology, University of Graz, [3]Genomics and Bioinformatics, Graz University of Technology, Graz, Austria, [4]Scientific Computing, [5]Biological Science, Florida State University, Tallahassee, United States

Abstract: With more than 1,700 described species, cichlid fishes are amongst the most species-rich vertebrate groups. Lake Tanganyika hosts the oldest, and therefore genetically and morphologically most diverse cichlid assemblage, representing a prime model for the study of adaptive radiation. Some haplochromine cichlids of this radiation left Lake Tanganyika again and, through rivers, colonized Lakes Malawi and Victoria founding their younger, but hyper-diverse cichlid species flocks. Phylogenetic reconstruction among the 250 species of Lake Tanganyika cichlids is challenging due to incomplete lineage sorting and repeated gene flow among distant lineages. Previous attempts have greatly contributed towards elucidating the cichlid family tree but uncertainties have remained, particularly some inter-tribal relationships. Here we present the most comprehensive attempt to disentangle the phylogenetic relationships using the anchored phylogenomic sequencing method (based on 950,000 sequenced nucleotides, 533 discrete loci derived from conserved regions and more variable flanking regions, and candidate loci of 149 species of East African cichlids). Concatenation and coalescent approaches reconstructed highly congruent and strongly supported phylogenies that are concordant with the current understanding of cichlid evolution, derived from previous smaller-scale studies. Patterns of tree incongruence suggest a significant role of ancestral inter- and intra-tribal gene flow (introgression or hybridisation) in shaping the extant Lake Tanganyika species flock. Analysis of molecular evolution in loci associated with key adaptive traits varied considerably possibly due to different selection pressures. Opsin genes, involved in cichlid vision, show evidence of strong positive selection associated with habitat choice or sexual selection.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-373

A diversity of diversifying selection on koala cytochrome P450 monooxygenase gene sequences

Catherine Grueber*

Abstract: Koala (*Phascolarctos cinereus*), an endemic Australian marsupial, feeds almost entirely on leaves from the *Eucalyptus* genus, a diet that would be toxic to most mammals. Sequencing the koala genome enables us to better understand the species' unique adaptations to this diet. The presented work forms part of this larger investigation (by researchers from the Koala Genome Consortium), by examining evidence of selection on cytochrome P450 monooxygenase (CYP) gene sequences across a multispecies alignment (N = 154 sequences, 33 from koala, as well as sequences from nine other species)*. I tested for diversifying selection in particular, utilising a mixed-effects model of episodic diversifying selection to reveal episodic selection: codons under positive selection in only a part of the tree, while under purifying selection elsewhere. The results were used to examine whether koala-specific sequences showed greater evidence of selection, at particular codons, relative to sequences from other species. Conserved regions of the alignment showed a strong tendency towards negative (purifying) selection, as would be expected for a functional protein. A total of 101 codons with high coverage showed evidence of episodic selection, with 60 codons showing significantly greater evidence for selection in koala-specific lineages than in other species. Furthermore, evidence of episodic selection in koala CYPs varied widely across genes, with 10 genes showing very low evidence of lineage-specific selection, and others much more variable. Collectively these results suggest that, despite a background of strong purifying selection generally, koala CYPs show evidence of diversifying selection: multiple genes are under different types of selection, and different codons appear to be under selection across genes. Together these findings have implications for our understanding of the evolution of toxin metabolism in mammals.
* Acknowledging Will Nash and Wilfried Haerty of the Earlham Institute for permitting use of their alignment.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

Molecular function limits divergent protein evolution on planetary timescales

Dennis Vitkup [1,*]

[1]Systems Biology, Columbia University, New York, United States

Abstract:

Protein homologues that share the same molecular function often retain significant sequence similarity over long evolutionary distances. Sequence comparisons of ancient proteins suggest that even after billions of years of independent evolution orthologous proteins are still diverging from each other. Although the evolution of protein function has been investigated in detail, whether the requirement to continuously maintain the same molecular function imposes an effective limit on divergent protein evolution of orthologs is not known. Here we address this question using several models of molecular evolution and sequence and structural analyses of multiple enzymes. Our results demonstrate that the mutual divergence rates of enzymes that maintain their molecular function have decreased more than 30 times during ~4 billion years of evolution. We find that the effective divergence limit is usually reached within 1-2 billion years of independent evolution and that the majority of orthologous enzymes are unlikely to diverge beyond ~25% sequence identity within planetary timescales. Notably, the divergence limit is usually reached well above the levels of detectible sequence homology. To better characterize the nature of the divergence constraint we experimentally characterize the relative growth rates of all amino acid substitutions in the *E. coli* protein folA. Interestingly, the probability that the average fitness effects of mutations at a protein site in *E. coli* correlates approximately linearly in two different proteins with the probability that the identical amino acids will occupy the site at the divergence limit. Our structural protein analysis shows that the divergence limit depends on the spatial proximity to the active site residues. Although the divergence limit is on average higher for sites that are closer to catalytic residues, the requirement to continuously maintain molecular function significantly constrains all protein regions. Finally, the present study allows us to distinguish, on the global scale, the convergent and divergent evolution of enzymes with the same molecular function.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-376

Diversity and evolution of leech anticoagulants

Michael Tessler [1,*], Sebastian Kvist [2], Mark Siddall [3]

[1]Richard Gilder Graduate School, American Museum of Natural History, New York, United States, [2]Royal Ontario Museum, Toronto, Canada, [3]Invertebrate Zoology, American Museum of Natural History, New York, United States

Abstract: Anticoagulants are critical for effective bloodfeeding in leeches, and, accordingly, leeches produce a pharmacopeia of powerful anticoagulants. This aspect of leech biology has led to the use of leeches in modern, authoritative medicine such as for reconstructive surgery. But despite their medical importance, only a small portion of leech diversity has been screened for anticoagulants. Even fewer studies have compared the evolutionary selection pressures on these proteins. Additionally, it is clear from genomic work that at least one non-bloodfeeding leech has retained anticoagulant genes. Here we sequenced transcriptomes for anticoagulants across a phylogenetically diverse group of leeches and leech relatives with a range of feeding and habitat preferences. With these data, we screened for anticoagulants and assessed their evolutionary selection and reconstructed their phylogenetic relationships. We found a broad array of anticoagulants, including proteins described from a number of animals that are venomous (e.g., snaclec from vipers) or bloodfeeding (e.g., apyrase from ticks). Our findings are the first to show that non-bloodfeeding leeches transcribe a diversity of anticoagulants from their salivary tissue. Positive selection was found for many codons and phylogenetic branches. The results were varied with a number of species under positive selection. Furthermore, a number of long branches were found in phylogenetic reconstructions, relating to the high levels of divergence seen in these genes.

Expanded summary*: Anticoagulants have been invaluable to medicine, pest control (e.g., rodenticides), and our understanding of disease vectors – more recently they have emerged as fascinatingly complex models in evolutionary studies (Mans et al. 2002; Rokyta et al. 2011; Song et al. 2011; Siddall et al. 2011; Low et al. 2013). For bloodfeeding leeches, anticoagulants are critical for survival and retaining mobility while digesting (Sawyer 1986). Leech anticoagulants still receive attention for medical use, but few studies have been conducted on the diversity or evolutionary selective pressures on these proteins (Siddall et al. 2011; Kvist et al. 2013). Recent advances in next-generation sequencing (NGS) of mRNA expressed sequenced tags (ESTs) now allow for rapid assaying of anticoagulants transcribed in leech salivary tissue (Kvist et al. 2014). The ESTs produced in these assays can then be leveraged toward producing phylogenomic reconstructions. Yet, at present, the closest relatives (Siddall et al. 2001) of leeches – Acanthobdellida, Branchiobdellida, and Lumbriculida – are critically lacking EST data, with non-salivary EST data only available from distant relatives (e.g., Haplotaxida on GenBank). My dissertation exploits NGS sequencing of ESTs from all major leech lineages and leech relatives to 1) resolve difficult to establish phylogenetic relationships between these annelids and within leeches; 2) determine anticoagulant diversity within leeches and their relatives, and establish the origin of these proteins in annelid evolution; and 3) test hypotheses on the evolution of anticoagulant diversity, including whether generalist leeches possess more

anticoagulants than specialists and whether anticoagulants in non-bloodfeeding leeches are being selected for are drifting from functionality.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-426

DETECTION OF REGIONAL VARIATION IN SELECTION INTENSITY WITHIN PROTEIN-CODING GENES USING DNA SEQUENCE POLYMORPHISM AND DIVERGENCE DATA

Zi-Ming Zhao [1], Michael Campbell [2], Ning Li [3], Daniel Lee [1], Zhang Zhang [4], Jeffrey Townsend [1,*]

[1]Yale University, New Haven, [2]Howard University, Washington, DC, [3]Novartis, Boston, United States, [4]Beijing Institute of Genomics, Beijing, China

Poster: Numerous approaches have been developed to infer natural selection based on the comparison of polymorphism within species and divergence between species. These methods are especially powerful for the detection of uniform selection operating across a gene. However, empirical analyses have demonstrated that regions of protein-coding genes exhibiting clusters of amino acid substitutions are subject to different levels of selection relative to other regions of the same gene. To quantify this heterogeneity of selection within coding sequences, we developed Model Averaged Site Selection via Poisson Random Field (MASS-PRF). MASS-PRF identifies an ensemble of intragenic clustering models for polymorphic and divergent sites. This ensemble of models is used within the Poisson Random Field (PRF) framework to estimate selection intensity on a site-by-site basis. Using simulations, we demonstrate that MASS-PRF has high power to detect clusters of amino acid variants in small genic regions and to reliably estimate the probability of a variant occurring at each nucleotide site in sequence data. We applied MASS-PRF to human gene polymorphism derived from the 1000 Genomes Project and divergence data from the common chimpanzee. Based on this analysis, we discovered striking regional variation in selection intensity, indicative of positive or negative selection, in well-defined domains of genes associated with neurological processing, immunity, and reproduction. We suggest that amino acid-altering substitutions within these regions likely are or have been selectively advantageous in the human lineage, playing important roles in protein function.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-397

Evolutionary Processes Underlying the Ecological Success of the Diaforarchaea, an Emerging Superclass of the Euryarchaeota

Julie Perreau*, Filipa Marques, Guillaume Borrel, Simonetta Gribaldo

Abstract: The genomic revolution has revealed a large number of new lineages in the Archaea, providing unprecedented opportunities to study the processes that have shaped current archaeal diversity. The Diaforarchaea (διάφορα, diáfora, miscellaneous), a recently defined superclass of the Euryarchaeota, are a particularly suitable group for addressing this issue. They include members that have adapted to life in a wide range of contrasted environments, from thermoacidophilic drainage sites and solfataric fields to deep-sea hydrothermal vents, open ocean water, and the intestinal tracts of termites, ruminants, and humans. With forty recently available genomes from culture, metagenomics, and single cell sequencing, we obtained a robust phylogenetic tree of the Diaforarchaea, based on multiple markers, and clarified the evolutionary relationships among their main lineages. Using a probabilistic approach, we performed ancestral reconstruction and mapped gene gain and loss events across the phylogeny in search of markers for adaptation to different environments. We defined gains as arising from either horizontal gene transfer (HGT) or innovation. Detailed phylogenetic analysis of potential HGT revealed that this phenomenon is a dominant evolutionary process that accompanied Diaforarchaea adaptation processes, with donors belonging to all three domains of life. As an example, we reveal an uncommon case of HGT from eukaryotes involving ribonucleotide reductase, which allowed a lineage of the Diaforarchaea to conquer oxygen-rich ocean waters. Overall, our analyses provide a robust framework for future studies focused on the role of HGT and innovation in adaptation processes of this widely distributed and ecologically important archaeal clade.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-405

Runs of homozygosity and their utility in conservation and evolutionary studies

Anna Brüniche-Olsen [1,*], J. Andrew DeWoody [1 2]

[1]Department of Forestry & Natural Resources , [2]Department of Biological Sciences, Purdue University, West Lafayette, United States

Abstract: Runs of homozygosity (ROH) are continuous stretches of autozygous nucleotides present in an individual's genome. They develop when consanguineous matings cause haplotypes that are identical-by-decent to be passed on to the offspring. Over time, recombination breaks down ROHs such that the prevalence and size of ROHs varies with demographic history. Thus, ROH lengths distributions can be used in conservation to quantify inbreeding and trace population size reductions. We used published whole genome sequences to characterize inbreeding and ROH patterns between i) domesticated and wild mammals; ii) threatened and non-threatened wild mammals; and iii) temperate and tropical mammal species. Our results provide a comparative genomic overview of inbreeding and may serve as reference for future studies of mammalian evolution and conservation.

Expanded summary*: Background

Inbreeding leads to loss of adaptation potential and increased extinction risk. A central aim in conservation programs is therefore to maintain maximum genetic diversity and avoid mating among relatives [1]. Inbreeding is prevalent in populations with small effective population size—e.g., endangered species, livestock, and isolated populations—where consanguineous mating causes haplotypes that is identical-by-decent (IBD) to be passed on to the offspring, resulting in runs of homozygosity (ROH).

ROHs are continuous stretches of homozygote genotypes present in an individual's genome [2]. Long ROHs are most likely IBD whereas shorter ROHs may arise due to chance. Over time recombination breaks down ROHs, thus the prevalence and size of ROHs varies with demographic history, and ROH lengths distributions can be used to trace population size reductions and past inbreeding [3].

Many conservation studies have compared natural populations to evaluate loss of genetic diversity and inbreeding levels using panels of microsatellites or a few hundred SNPs [4], but in order to capture accurate genome-wide heterozygosity and inbreeding levels, marker density needs to be substantially higher [5]. Next generation sequencing is likely to improve our understanding of ROHs and their usefulness as a tool for measuring inbreeding as genome-wide sequence data enables accurate detection of individual autozygosity [2].

This project aims at investigating differences in genome-wide inbreeding levels across a range of mammal species. The aim is to measure extend of autozygosity in genomes of mammals with published genome data in order to obtain ROH size distribution for each species. We will determine differences in inbreeding levels among groups of species—i.e., wildlife and domesticated species, endangered and non-endangered species—and from the ROH distributions estimate when the inbreeding most likely occurred. Our results may serve as a reference for researchers investigating inbreeding in both domesticated and wildlife species.

Objectives

In this project we use published genomes from ~130 mammal species to characterize levels of inbreeding and ROH patterns. The objectives was to determine if there are differences among groups of species i.e.: i) domesticated and wildlife species, ii) threatened and non-threatened species, and iii) temporal and tropic species.

Methods

We use whole genome sequences and short read archive (SRA) data from NCBI to quantify inbreeding and ROH in ~130 published mammal species. Briefly, for each species we trimmed low quality bases and adaptor sequences [6], quality checked the cleaned reads [7], mapped them to the genome [8], removed duplicates and called SNPs [9],and identified ROHs [10]. For each species we estimated five different parameters: i) nucleotide diversity across the entire genome (π_{genome}); ii) nucleotide diversity outside ROHs (π_{noROH}); iii) number of ROHs in each genome; iv) the mean length of ROHs; and v) the inbreeding coefficient F_{ROH}, the overall proportion of the genome contained in ROHs.

Significance

Our results provide comparative overview of inbreeding levels and timing of population size reductions across mammal species. The majority of genome-wide ROH estimates have been done in humans and livestock species, thus our study will significantly extend the list of species, and may serve as reference for other studies investigating inbreeding from whole genomes.

References

1. Frankham R: **Genetics and extinction**. *Biol Conserv* 2005, **126**(2):131-140.
2. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A: **Runs of homozygosity in European populations**. *The American Journal of Human Genetics* 2008, **83**(3):359-372.
3. Kardos M, Qvarnström A, Ellegren H: **Inferring Individual Inbreeding and Demographic History from Segments of Identity by Descent in *Ficedula* Flycatcher Genome Sequences**. *Genetics* 2017.
4. Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW: **Genomics advances the study of inbreeding depression in the wild**. *Evolutionary applications* 2016, **9**(10):1205-1218.
5. Curik I, Ferenčaković M, Sölkner J: **Inbreeding and runs of homozygosity: A possible solution to an old problem**. *Livestock Science* 2014, **166**:26-34.
6. Schubert M, Lindgreen S, Orlando L: **AdapterRemoval v2: rapid adapter trimming, identification, and read merging**. *BMC Research Notes* 2016, **9**(1):88.
7. **FastQC A Quality Control tool for High Throughput Sequence Data** [http://www.bioinformatics.babraham.ac.uk/projects/fastqc/]
8. Li H, Durbin R: **Fast and accurate long-read alignment with Burrows-Wheeler transform**. *Bioinformatics* 2010, **26**:589 - 595.
9. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al*: **The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data**. *Genome Res* 2010, **20**.
10. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, Sklar P, de Bakker P, Daly M *et al*: **PLINK: a tool set for whole-genome association and population-based linkage analyses**. *Am J Hum Genet* 2007, **81**:559 - 575.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-350

Subfunctionalization of SRPK—a new Y-linked gene family in the Drosophila simulans clade

Ching-Ho Chang [1,*], Amanda Larracuente [1]

[1]Department of Biology, U of Rochester, Rochester, United States

Abstract: Non-recombining Y chromosomes in *Drosophila* species are typically gene-poor compared to X chromosomes, and acquire most of their genes from autosomes. Most of what we know about *Drosophila* Y chromosome dynamics is based on studies of the ~20 Y-linked genes in *D. melanogaster* over long evolutionary time scales. We used long-read sequencing from Pacific Biosciences in three *simulans* clade species to assemble the Y chromosomes. Our Y-linked contig N50s range from 500 Kb to 1.5 Mb. We reveal at least 4 duplications from elsewhere in the genome specific to the *simulans* clade Y chromosomes. Most notable of these is the Y-linked acquisition and amplification of *serine-arginine protein kinase* (*SRPK*), shared in the *simulans* clade. *SRPK* is ubiquitously expressed in *Drosophila*, with roles in both oogenesis and spermatogenesis, suggesting a possible history of intralocus sexual conflict. One resolution of this conflict is through subfunctionalization following duplication to the Y chromosome —*SRPK-Y*s retained exons from a testis-specific isoform deleted in the autosomal copies. These *SRPK-Y*s evolve rapidly at the sequence and expression levels. *SRPK-Y*s have 3-fold higher protein evolution rate and are 4-fold overexpressed in the testes of *D. simulans* compared to their parental copy in *D. melanogaster*. These data suggest that sexual antagonism may have played a role in the evolution of *SRPK-Y*s. Interestingly, we find copy number variation in all known Y-linked genes, including *SRPK-Y*s. We hypothesize that gene duplications may accelerate Y chromosome evolution in *Drosophila* species.

Expanded summary*: Y chromosomes are typically degenerated—they contain few genes but are rich in repeats. Despite these features, Y chromosomes can be under strong selection and affect many traits. Repetitive sequences make Y chromosomes difficult to study. We use long read single molecule sequencing (PacBio) to circumvent some problems with repeats and assemble Y chromosomes. I obtained deep (~100X) PacBio data for 3 *simulans* clade species. My Y chromosome assemblies in these species provide an opportunity to study Y chromosome function and evolution over short evolutionary timescales—these species diverged only 240 Kya. Using these assemblies, I discovered the rapid evolution of Y-linked gene copy number, intron size, repeat content and gene order. I also studied gene traffic between the Y and other chromosomes to understand the forces driving Y-linked gene acquisition in *Drosophila*. I identified 4 independent duplications specific to the *simulans* clade Y chromosomes. One new Y-linked gene family is particularly interesting: s*erine-arginine protein kinase* (*SRPK*), which duplicated to the Y chromosome of the *simulans* clade ancestor and now shows evidence of potential intralocus sexual conflict. After the duplication, both autosomal and Y-linked copies (*SRPK-Y*) subfunctionalized through deletions—*SRPK-Y*s retained exons from a testis-specific isoform deleted in the autosomal copies, and lost ubiquitously expressed isoforms. *SRPK-Y*s are present in multiple copies on the Y chromosomes of each species, and evolve rapidly at the sequence and expression levels. I aim to understand the functional divergence between *SRPK* and *SRPK-Y* and their roles in spermatogenesis. My study not only addresses the evolution of Y-linked genes but further suggests that subfunctionalization of *SRPK* and *SRPK-Y*s may contribute hybrid incompatibilities in of *D. melanogaster* and *simulans* clade.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-399

Molecular dating when clades evolve with different rate models

Qiqing Tao [1,*], Beatriz Mello [2], Koichiro Temura [3], Sudhir Kumar [1]

[1]Biology, Temple Univeristy, Philadelphia, United States, [2]Genetics, Federal University of Rio de Janeiro, Rio, Brazil, [3]Biological Sciences, Tokoyo Metropolitan University, Tokoyo, Japan

Abstract: Evolutionary rates can vary extensively in a large phylogeny, which necessitates the use of methods that allow for rate variation among lineages when estimating divergence times. Bayesian methods incorporate specific rate models, such as the lognormal distribution, to account for rate variation among lineages throughout the tree. However, the patterns of rate variation may differ among clades in large phylogenies, which would make the use of a single rate model for the whole tree inappropriate when dating divergences. Indeed, Bayesian methods that use a single rate model have been shown to produce biased divergence times when there are two clade-specific rate models in the tree. Here we show that the RelTime method, which accounts for rate variation among lineages but does not require selection of a rate model a *priori*, performs well in estimating divergence time when rate models are different between clades. We suggest that estimates from RelTime and Bayesian methods be compared to detect potential biasing effects of the existence of multiple rate models in a phylogeny.

Expanded summary*: Time-calibrated phylogenies (timetrees) are important to understanding the temporal patterns of life evolution and biodiversity. The knowledge of precise and accurate speciation time is needed to discover the relationship between the geological events and environmental changes. Among various molecular dating methods, Bayesian methods are widely used in time estimation because they enable us to incorporate the pattern of rate variation among branches and fossil calibrations as priors. Indeed, Bayesian methods can provide reliable timescales when the priors are correctly specified. However, the information of the rate variation pattern in a phylogeny, the speciation process, and the fossilization are rarely known reliably in practice. It makes the choice of priors difficult, and the incorrect specification of priors in Bayesian methods can lead to biased results. Thus, it is urgent to develop methods that can aid in choosing priors objectively to improve accuracy and precision of the current Bayesian estimates. In my research, I am developing empirical Bayes approaches to select the best-fit rate model and evaluate calibrations. Empirical Bayes approach is different in that the priors are determined using the data beforehand. Therefore, more informative priors are used in empirical Bayes approach, which provides an opportunity for more accurate and precise estimates. My theoretical and data analysis research will greatly improve the accuracy and precision of the timetree estimation and enhance the power of testing biological hypotheses that directly or indirectly involve the timing of evolutionary events. To make my method widely accessible, I will develop software and educational resources for teachers, students, and researchers. We will develop open-source software with optional web-based implementation for researchers and students in diverse fields, including evolution, ecology, and genomics. Our methodological advances made accessible by software will enable the widespread and efficient use of biological data in discovery and learning across diverse disciplines.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-398

Impact of effective population size on sex chromosome evolution among closely related Silene species

Aline Muyle*, Niklaus Zemp, Alexander Widmer, Gabriel Marais

Abstract: The effective population size (N_e) is a major force affecting genome architecture. It has been suggested that many of the features that differ among sex chromosome systems may be related to differences in N_e among species. The non-recombining Y is indeed expected to degenerate faster with smaller N_e, which should lead to faster Y gene loss and faster decrease in Y expression. Also, the efficacy of selection for dosage compensation, a mechanism that balances male and female expression in spite of Y degeneration, should be decreased with small N_e. Here we took advantage of the extreme variation in geographic distributions within the *Melandrium* section of *Silene* (*S. latifolia*, *S. dioica*, *S. heuffelli*, *S. marizii* and *S. diclinis*) to test these predictions. First, we used a population transcriptomic approach to estimate silent genetic diversity in the five closely related species using 3 to 17 individuals coming from different populations, which revealed contrasted N_e among species, as expected from their diverse geographic distributions. Second, we obtained transcriptomes of individuals from crosses (parents and progeny) to identify the genes located on the sex chromosomes with the SEX-DETector program in each of the five species. We then estimated the rate of Y gene loss among species, as well as the decrease in Y expression and the intensity of dosage compensation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS8

Ancient Adaptive landscape of the human coding genome

Sudhir Kumar [1,*]

[1]iGEM, Biology, Temple University, Philadelphia, United States

Abstract: Hundreds of thousands of missense mutations that alter the amino acid sequence of a protein have been discovered in the human genome. But, only a handful of these variants are known to be adaptive, which implies that adaptation through protein sequence change is an extremely rare phenomenon in the human evolution. Alternatively, existing methods may lack the power to pinpoint adaptive variation. We have developed and applied an Evolutionary Probability test, which utilizes interspecific differences to generate neutral expectations of the occurrence of a given amino acid residue in the human genome and discovers candidate adaptive polymorphisms (CAPs) through the discordance of allelic evolutionary probabilities and observed frequencies in human populations. We report more than 18,000 missense CAPs, of which more than 2,000 are functionally implicated in available genotype-phenotype association data at a stringent $P < 10^{-8}$. Moreover, CAPs predominantly originated in the ancestors of modern humans and represent ancestral standing variation. Thus, our results reveal a human coding genome in which thousands of protein polymorphisms bear signatures of ancient adaptation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-400

Aging at the Molecular Level in Wild Baboons

Jordan Anderson [1,*], Amanda Lea [2], Susan Alberts [1 2 3], Elizabeth Archie [3 4], Jenny Tung [1 2 3 5]

[1]Department of Evolutionary Anthropology, [2]Department of Biology, Duke University, Durham, United States, [3]Institute of Primate Research, Nairobi, Kenya, [4]Department of Biological Sciences, University of Notre Dame, Notre Dame, [5]Duke Population Research Institute, Duke University, Durham, United States

Abstract: Physiological decline with age is a general feature of most organisms. However, even within species not all individuals experience this decline at the same rate. In animals, blood-based molecular markers have emerged as a useful tool for measuring this variation. Nevertheless, it remains unclear whether these assays reflect overall condition or instead capture trade-offs in investment between systems (e.g. investment in immune function at the cost of growth). To address this gap, we analyzed genome-wide gene expression data generated from whole blood from an intensively studied wild baboon population (RNA-seq data for n=63 known-age individuals). As expected, we observed pervasive effects of age on gene expression levels (n=1058 age-associated genes at a 10% false discovery rate), enriched for age-associated biological processes such as cellular communication and immune defense ($p<5e-5$). Based on the gene expression data alone, an elastic net regression model was able to predict chronological age with good accuracy ($r^2=0.58$, $p<5e-10$ using leave-one-out cross-validation). Residuals calculated from regressing predicted age on chronological age were positively correlated with body mass index (BMI: $r^2=0.18$, $p=0.0048$). Thus, individuals predicted to be younger than their chronological age (by their gene expression profile) tended to have lower BMI, an indicator of poorer body condition in this population. These results suggest that individuals able to defer physiological senescence in immune cells may bear costs in other systems, consistent with the predictions of classical life history theory.

Expanded summary*: Physiological decline with age is a general feature of most organisms. However, even within species not all individuals experience this decline at the same rate. In animals, blood-based molecular markers have emerged as a useful tool for measuring this variation. Nevertheless, it remains unclear whether these assays reflect overall condition or instead capture trade-offs in investment between systems (e.g. investment in immune function at the cost of growth). To address this gap, we analyzed genome-wide gene expression data generated from whole blood from an intensively studied wild baboon population living in the Amboseli ecosystem of Kenya (RNA-seq data for n=63 known-age individuals).

As expected, we observed pervasive effects of age on gene expression levels (n=1058 age-associated genes at a 10% false discovery rate), enriched for age-associated biological processes such as cellular communication and immune defense ($p<5e-5$). Based on the gene expression data alone, an elastic net regression model was able to predict chronological age with good accuracy ($r^2=0.58$, $p<5e-10$ using leave-one-out cross-validation). Residuals calculated from regressing predicted age on chronological age were positively correlated with body mass index (BMI: $r^2=0.18$, $p=0.0048$). Thus, individuals predicted to be younger than their chronological age (by their gene expression profile) tended to have lower BMI, an indicator of poorer body condition in this population. These results suggest that individuals able to defer physiological senescence in immune cells may bear costs in other systems, consistent with the predictions of classical life history theory. Aging remains the leading cause of morbidity and mortality in humans, yet we still know surprisingly little about how other organisms cope with this pervasive pressure. By demonstrating the potential for trade-offs between physiological systems in response to senescence, we hope to promote future investigation into how the cost of aging may be distributed across biological systems in other animals.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-366

Optimal estimation of FST for detecting positive selection from SNPs under sample size bias, missing values, and low frequency variants

Songeun Lee*, Yuseob Kim [1]

[1]Ewha womans university, Seoul, Korea, Republic Of

Abstract: Wright's F_{ST} is widely used as a robust signature for detecting positive selection, for example a complete or incomplete selective sweep in a local population. Given the basic formula, $F_{ST} = 1 - H_W/H_T$ where H_W is mean within-population diversity and H_T is total population diversity, various methods for estimating F_{ST} from DNA sequence were proposed. However, when the sizes of two populations are not similar, simple weighting by sample sizes makes both H_W and H_T predominantly determined by the population of larger sample size, which yields unrealistically small (or negative) value of F_{ST}. In addition, it is not clear how robust the calculation of F_{ST} is in the presence of missing data (missing base calls) in NGS-based genomic data. Furthermore, as the upper bound of F_{ST} is an increasing function of minor allele frequency (MAF), one needs to find a way to correct the effect of MAF on the statistic. To address these problems, we devised three ways of estimating F_{ST}. First, H_W and H_T are calculated from mean pairwise sequence differences but after random subsampling to make the sample sizes of two populations similar. Second, we averaged F_{ST} calculated per individual sites. Here allele frequencies are obtained by giving equal weights to populations in calculating H_T. Third, the above method is modified to obtain heterozygosity first from each population and then from the total population. Analysis of simulated data indicates that the third method, after removing sites with low MAF, is optimal in the presence of missing data.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-403

Patterns of mito-genomic evolution in rattlesnakes: do disparate phylogenetic signals indicate historical recombination?

Jeffrey Streicher [1,*], Longson Pang [2], Daniel Dashevsky [3], Matthew Fujita [4], Sarah Schaack [3], Jesse Meik [5]

[1]Department of Life Sciences, The Natural History Museum, [2]University College London, London, United Kingdom, [3]Department of Biology, Reed College, Portland, [4]Department of Biology, The University of Texas at Arlington, Arlington, [5]Biological Sciences, Tarleton State University, Stephenville, United States

Abstract: Animal mitochondria are inherited as a single linkage unit and typically do not recombine. However, mitochondrial recombination has been observed in some species, albeit rarely. As part of a study on geographic variation in mitochondria, we sequenced whole mitochondrial genomes from four species of closely related rattlesnakes (Genus *Crotalus*). Using genomes from 14 individuals we generated phylogenies in two ways: (1) by using an alignment of the whole-genome and (2) by using individual alignments of 13 protein-coding genes and 2 ribosomal genes. Contrary to our expectation that the protein-coding gene trees would be similar, we observed two distinctive phylogenies across the genome. One topology is similar to the whole-genome topology, whereas the other features key differences regarding the placement of two taxa. We also observed that the four genes (ND1, ND2, CO1, CO2) featuring the key differences displayed significantly more variance in branch-specific substitution rates than the other genes. Intriguingly, the disparate phylogenetic signals were not randomly distributed on the mitochondrial genome – they clustered according to gene order. We explore various explanations for this unexpected pattern including interspecific recombination, laboratory methodology, and selection.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-362

Ancient functional optimization of molecular dynamics of nucleic acids may have led to the triplet-based organization of the genetic code

Gregory Babbitt*, Erin Coppola [1], Jamie Mortensen [1], Andre Hudson [2] and RITgroup

[1]Biomedical Engineering, [2]T.H. Gosnell School of life Sciences, Rochester Institute Of Technology, Rochester, United States

Abstract: Since the elucidation of the genetic code almost 50 years ago, many aspects of its non-random codon organization remain only partly resolved. Here we investigate the recent proposal that degeneracy in the triplet-based genetic code, in addition to facilitating translational efficiency via codon optimization, also functions to accommodate change in the molecular dynamics of nucleic acids while not interfering with protein synthesis. We analyzed synonymous and nonsynonymous impacts in 13823 identically degenerate alternative codon reorganizations, defined by codon transitions on 3200 massively parallel molecular dynamic simulations on implicitly solvated DNA and RNA structures containing all 64 codons. When compared to randomized alternative codes, we find that the genetic code minimizes mutational impacts in the correlated dynamics of carbon backbone at nonsynonymous (protein-altering) sites, while maximizing impacts of mutation at synonymous (protein-silent) sites. Identical patterns were also confirmed using empirically-derived scoring metrics of DNA flexibility and RNA stability. This divergent optimization of the nucleic acid polymer dynamics of codons implies that the code may have resulted from an adaptive functional expansion event over 3.8 billion years ago, enabling a primordial doublet code, pre-adapted towards accurate and efficient translation, to also multiplex biophysical information at synonymous sites without interfering with protein synthesis.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-352

Bio42 - Graphing Genetic Flow

Martin Rusilowicz [1,*], James McInerney [1]

[1]School of Biological Sciences, The University of Manchester, Manchester, United Kingdom

Abstract: Evolutionary history is usually studied using phylogenetic trees. However, because of the ubiquity of recombination and horizontal gene transfer, modelling evolution according to a phylogenetic tree is not always appropriate. Modern, high-throughput genome sequencing technologies have resulted in the generation of massive datasets encompassing thousands of different organisms. In addition, common analytical techniques such as all-vs-all BLAST searches generate even more data. Despite being well suited for a system-biological approach to analysis, the storage and retrieval of such data is largely governed by reductionist principles, segregating data by predefined classifications such as taxonomy or disease-class. In this poster we present an on-going study into the creation of a large genomic database driven from a graph-theoretical perspective. The relationships in this database can include statements of homology, as judged by BLAST or other similar approaches. In contrast to existing "relational" or "flat" databases this permits not only the exploration of entities in relation to each other, but facilitates the study of the system as a whole. Further, due to the wealth of existing information, we include the computational tools and software necessary for visualisation, analysis and importation of these data.

DATA AND SOFTWARE ARE FREE AND OPEN-SOURCE
THIS PROJECT IS FUNDED BY BBSRC / GRANT NUMBER BB/N018044/1

Expanded summary*: My research focuses on the study of introgression and horizontal gene transfer (HGT) events. To this end a systems-biology approach is taken, which seeks to make use of the vast amounts of information now available in the genomics field.

In contrast to taking a tree-like, or phylogenetic approach to evolution, my research at the McInerney lab seeks to develop "genetic networks" -- arguably a systems-biology approach to a systems-biology problem. These networks present the information flow between individual organisms, genomes, and meta-genomes as a series of edges and nodes on a k-partite graph, including, but not limited to, known taxonomic information, gene ontology annotations and BLAST query results.

This is significant on several levels. Considering the system as a graph permits a deeper understanding of evolutionary history, allowing both horizontal and vertical evolutionary events to be identified and mapped. Unlike time-based trees, the flow of genetic information can be followed through aspects such as space/locale and species. Second, whilst some genomes can be connected through known pyogenic events (X evolves to Y) in many cases this is either unknown or the calculation is computationally prohibitive. In this case genes can related through use of common analysis tools such as BLAST (X is similar to Y). Third, this technique permits the multitude of existing mathematical network analytical methods to be applied to large-scale gene networks, such as community detection and clustering. Finally, non-rigid graph structure is highly receptive to new data, which will either strengthen or weaken existing relationships between genes in the graph.

Whilst there are many existing studies that consider evolution in a network-wise manner there remains no simple method of generating and storing such networks. A second aspect of my research is therefore to make such methods readily available to evolutionary researchers. Our small sample-dataset consists of 100 thousand gene sequences (nodes) and 10 million edges (nodes). In addition to providing a front-end to this large-scale data storage, allowing for the importation and exportation of data, I am researching new and existing methods of visualising such information.

The database and all software will be released under a free and open-source license.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS4

Divergent evolution of olfactory and taste receptor repertoire in New World monkeys with diverse color vision types and feeding habits

Shoji Kawamura [1,*], Takumi Naoi [1], Masahiro Hayashi [1], Ryuichi Ashino [1], Yoshihito Niimura [2 3], Kazushige Touhara [2 3], Carrie Veilleux [4], Eva Garrett [5 6], Amanda Melin [5 6]

[1]Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, [2]Graduate School of Agricultural and Life Sciences, The University of Tokyo, [3]ERATO Touhara Chemosensory Signal Project, JST, Tokyo, Japan, [4]Department of Anthropology, University of Texas at Austin, Austin, United States, [5]Department of Anthropology & Archaeology, [6]Cumming School of Medicine, University of Calgary, Calgary, Canada

Abstract: Primates are generally regarded as vision-oriented animals, for which other senses, especially olfaction, are less important. However, this notion is questioned by recent studies and requires further investigation. New World monkeys (NWMs) are particularly suitable for understanding the interplay of different senses in evolutionary and ecological contexts because of their diversity in color vision and diets. Here, by employing targeted capture and massively parallel (Next Generation) sequencing methods, we studied the entire gene repertoire of olfactory receptors (ORs) as well as bitter (TAS2Rs) and umami/sweet (TAS1Rs) taste receptors from 6 species of NWMs in three Families (Cebidae: tufted capuchin, saddle-back tamarin, Azara's owl monkey; Atelidae: black-handed spider monkey, mantled howler monkey; Pitheciidae: dusky titi), with diverse color vision types and feeding habits. Contrary to a sensory “trade-off” prediction, the proportion of OR pseudogenes was not highest in the howler monkey, the sole routine trichromatic genus in NWMs, compared to other species with dichromacy/trichromacy polymorphism. The number of intact and defective OR genes appeared to differ only slightly among taxa, whereas genetic composition differed considerably among species by repeated gain and loss of OR genes throughout phylogeny. Bitter taste TAS2Rs were generally higher in nucleotide polymorphism than neutral references, many of which were nonsynonymous, implying diversifying evolution in bitter sensation. In contrast, functional constraint on umami/sweet taste TAS1Rs appeared generally relaxed, with the umami taste TAS1R1 gene in tamarins pseudogenized. These results depict a feature of active turnaround of gene contents in the evolution of primate chemical sensors.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-358

Modeling metapopulation persistence under changing climate

Mikhail V. Matz [1,*]

[1]Integrative Biology, University of Texas at Austin, Austin, TX, United States

Abstract: I have developed a multi-QTL model of metapopulation adaptation where individual populations adapt to their local environmental optima while exchanging migrants. The model is implemented in SLiM and can be easily adjusted to incorporate a variety of demographic, genetic and environmental scenarios. Here, I used the model to see which factors make a metapopulation more resilient to climate change. Predictably, longer metapopulation persistence was observed with larger number of QTLs, larger QTL effect size, larger population sizes, higher phenotypic plasticity, and migration rates that are not too high and not too low. One counter-intuitive finding was that higher heritability resulted in faster extinction. Although lower heritability reduced the maximum fitness that a metapopulation could attain under stable conditions, it substantially increased its chance of persistence under changing climate since it promoted higher standing genetic variation. Another non-trivial result was that random environmental fluctuations led to dramatic fitness fluctuations under changing conditions but not under stable conditions. Most importantly, I find that under a range of realistic scenarios a metapopulation is able to successfully adapt to changing climate for over 100 generations based solely on standing genetic variation existing among locally adapted populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-355

Conflicting evolutionary histories of the mitochondrial and nuclear genomes in New World Myotis

Roy Platt II, Brant Faircloth, Kevin Sullivan, Troy Kieran, Travis Glenn, Micheal Vandewege, Tom Lee, Robert Baker, Richard Stevens, David Ray [1,*]

[1]Biological Sciences, Texas Tech University, Lubbock, United States

Abstract: The diversification of *Myotis* into more than 100 species in just a few million years is one of the most extensive mammalian radiations available for study. Efforts to understand relationships have primarily utilized mitochondrial markers. The few studies to use nuclear markers resulted in trees that lack resolution. Thus, our understanding of relationships within *Myotis* is biased towards a set of markers that may not reflect the true species tree. To resolve this we sequenced the full mitochondrial genomes of 37 representative *Myotis*, primarily from the New World, in conjunction with targeted sequencing of more than five thousand ultraconserved elements (UCEs). We explored various concatenation and summary methods and combinations of markers based on informativeness or levels of missing data. Of the 293 phylogenies generated from the nuclear data, all are significantly different from phylogenies inferred from the mitochondrial genomes. Several factors can drive such conflict including horizontal gene transfer, hybridization, and lineage sorting. Quartet frequencies indicate that around half of all UCE loci conflict with the estimated species tree. This suggests that *Myotis* genomes experienced massive amounts of incomplete lineage sorting, likely during the early stages of the radiation. Mitochondrial genomes, because they evolve as single loci, are likely to have been incompletely sorted along with nuclear loci, meaning that phylogenies inferred using this locus probably do not reflect the true species tree. Based on these results, we re-examine the evolutionary history of *Myotis* to better understand the phenomena driving the unique nuclear, mitochondrial, and biogeographic histories.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-363

Building your own home-brew GPU-accelerated molecular dynamic modeling workstation from computer gaming hardware

Gregory Babbitt*

Abstract: Molecular evolutionary studies have a long history of computer-based sequence analysis, more recently enriched with molecular structural studies. A molecular viewpoint that is still missing from the field is that of molecular dynamics; the way in which macromolecules interact and move. The ability to examine the evolution of molecular dynamics is of particular relevance to understanding DNA-protein and protein-protein interactions that underlie epigenetics and gene regulatory evolution. However, broad application of molecular dynamic simulation to large biological systems has traditionally been hindered by its massive computational expense (i.e. time steps must be made in femtoseconds on several parameters for all individual atoms in a system). The very recent advent of high end computer gaming graphics processors (GPUs) now makes it possible to build many thousands of computing cores into a relatively cheap desktop workstations. This now enables molecular dynamic simulation of biologically relevant phenomena without the need for supercomputing resources. Here, we demonstrate the process of building and running your own state-of the-art Linux-based GPU molecular modeling workstation based on AMBER16 and Chimera 1.11 software for less than $5000. We demonstrate a range of technical solutions for (A) solving typical Nvidia graphics incompatibility problems encountered on Linux, (B) installing and testing CUDA GPU programming language and development tools, and (C) installing and testing GPU accelerated AMBER16. We show some simple Perl scripting solutions for pipelining and scheduling jobs to designated graphics cards. We also offer some statistical advice for making molecular evolutionary inferences about molecular mechanics and dynamics.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-353

Genomics of hybridizing ground squirrels

Roy Platt [1,*], Michael Vandewge [1], Cody Thompson [2]

[1]Biological Sciences, Texas Tech University, Lubbock, [2]Museum of Zoology, University of Michigan, Ann Arbor, United States

Abstract: The opportunity to understand the process of speciation at the genomic level has increased due to the availability of reduced representation and high-throughput sequencing. Two species of squirrel, Rio Grande Ground Squirrel (*Ictidomys parvidens*) and Thirteen-lined Ground Squirrel (*I. tridecemlineatus*), hybridize in isolated populations within southeastern New Mexico and West Texas. The hybrid zone is a result of secondary contact due to changes in climate occurring within the past 100 years. We used double digestion restriction-site associated DNA sequencing (ddRADseq) to identify single nucleotide polymorphisms in hundreds of individuals, including known hybrids, to quantify genomic introgression and hybridization in *I. parvidens* and *tridecemlineatus*. Our results demonstrate frequent hybridization between *I. parvidens* and *tridecemlineatus* with substantial genomics introgression between species. In addition, we identified a number of F2 and backcross hybrids implying that postzygotic reproductive barriers, in the form of selection against hybrid phenotypes, is minimal. Despite high levels of gene flow between species, genomic regions of high F_{st} were identified in each species. It remains to be seen if there is a breakdown of reproductive barriers in the *I. parvidens/tridecemlineatus* species pair. Regions of high F_{st} may contain loci responsible for limiting introgression and determining the forces responsible for reproductive isolation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-359

Speciation in endemic freshwater gastropods from the Kaek River, Thailand

Nora Lentge-Maaß [1,*], Matthias Glaubrecht [1]

[1]Center of Natural History Hamburg (CeNak), Hamburg, Germany

Abstract: Ecologically driven sympatric speciation is currently accepted as an alternative explanation for speciation, after the traditional view of allopatric speciation as the usual mechanism long dominated. Lacustrine organisms, among them also gastropods, have proven to be ideal model organisms to study the diverse mechanisms involved in speciation and adaptive radiation. Most of these lacustrine models have been studied extensively and appeared to be rather complex, while radiations in a riverine environment are rare and remain poorly studied. The unique setting at the Kaek River in Thailand, that hosts about ten named species of *Brotia* snails, most of them endemic, provides a perfectly suitable model system to study fundamental mechanisms acting in the speciation process. In an integrative approach geometric morphometrics of the shell, morphological differentiation of the radula and life history data (i.e. viviparous reproductive strategies) are combined with ecological parameters in order to assess phenotype–environment correlations. This dataset will be supplemented by genetic data on mitochondrial divergence using COI and whole genome data using –for the first time in freshwater gastropods– RAD sequencing. Thus, our study combines morphologic traits with evolutionary genetics, enabling a comparative assessment of ecological vs. geographical factors as drivers in speciation.
Keywords: morphometrics; morphology; species delineation; RAD sequencing; viviparity; Cerithioidea; Pachychilidae

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-395

Stressful removal of a selfish element

Marina Rudan [1], Tobias Warnecke [2 3,*], Anita Krisko [1]

[1]Mediterranean Institute for Life Sciences, Split, Croatia, [2]Institute of Clinical Sciences, Imperial College London, [3]MRC London Institute of Medical Sciences, London, United Kingdom

Abstract: Self-splicing introns are mobile genetic elements that populate several highly conserved protein-coding genes in fungal and plant mitochondria. They have persisted in yeast populations over long periods of evolutionary time despite making no adaptive contribution to their host (as far as we know). In part, their persistence in the host genome has been attributed to their capacity to continuosly re-invade the population. In addition, it has been suggested that they have a minimal impact on host fitness given their ability to splice themselves out of their host mRNAs and thereby reconstitute functional reading frames. In this talk, I will challenge this notion by exposing both substantial physiological costs of self-splicing introns in *Saccharomyces cerevisiae.* In particular, I will show that, unexpectedly, removing self-splicing introns is stressful to the host, with effects on growth rate, life span, mitochondrial morphology and function. Based on a series of experiments to unravel the nature of costly intron removal, I will suggest that the persistence of *S. cerevisiae* self-splicing introns has been facilitated by an evolutionary lock-in event, whereby the host genome adapted to an initial invasion in a way that incidentally rendered subsequent intron loss deleterious. I will suggest more generally, and counterintuitively, that the long-term persistence of some mobile elements might be facilitated by adaptive changes in the host genome that ameliorate deleterious effects of mobile element activity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-392

Targeted sequencing of venom genes from cone snail genomes reveals coupling between dietary breadth and conotoxin diversity

Mark Phuong*

Abstract: Although venomous taxa provide an attractive system to study the genetic basis of adaptation and speciation, the slow pace of toxin gene discovery through traditional laboratory techniques (e.g., cDNA cloning) have limited their utility in the study of ecology and evolution. Here, I applied targeted sequencing techniques to selectively recover venom gene superfamilies and non-toxin loci from the genomes of 32 species of cone snails (family, Conidae), a hyper diverse group of carnivorous marine gastropods that capture their prey using a cocktail of neurotoxic proteins (conotoxins). I was able to successfully recover conotoxin gene superfamilies across all species sequenced in this study with high confidence (> 100X coverage). Through comparative genomic analyses of the recovered conotoxin loci, I provided evidence for several genetic factors shaping venom composition in cone snails, including positive selection, extensive gene turnover, and expression regulation. These results suggest that multiple molecular mechanisms make important contributions to the extensive variation documented in the venom phenotype. Using comparative phylogenetic methods, I found that while diet specificity did not predict patterns of conotoxin gene superfamily size evolution, dietary breadth was positively correlated with total conotoxin gene diversity. These results continue to emphasize the importance of dietary breadth in shaping venom evolution, an underappreciated ecological correlate in venom biology. Finally, the targeted sequencing technique demonstrated here has the potential to radically increase the pace at which venom gene families are sequenced and studied, reshaping our ability to understand the impact of genetic changes on ecologically relevant phenotypes and subsequent diversification.

Expanded summary*:

BACKGROUND: Understanding the molecular basis for adaptation and speciation is a central goal in evolutionary biology. Although large-scale comparative genomic studies have vastly increased our knowledge of the genetic changes associated with diversification, the link between genotype and ecologically relevant phenotypes frequently remains unclear. Often, the functional consequences of genetic patterns such as an excess of gene duplicates or regions under positive selection are unknown, limiting our ability to understand how genetic changes shape the evolutionary trajectory of species.

Animal venoms provide an excellent opportunity to study the interplay between genetics and adaptation because venom gene families are (a) well-characterized and (b) typically encode toxic proteins that have a direct impact on prey capture and survival. A fundamental challenge associated with the study of venom evolution is the inability to rapidly obtain sequences from venomous multi-gene families. Traditionally, venom genes were sequenced through cDNA cloning techniques, which can be labor intensive and time-consuming. Here, I improve upon this limitation by using a targeted sequencing approach to study the evolution of venom gene families across 32 species of cone snails (family, Conidae), a diverse group of venomous marine gastropods.

RESULTS: Through successful application of targeted sequencing techniques and analyses of the recovered conotoxin loci, I provide comprehensive support for the following genetic factors influencing the venom phenotype:

(a) positive selection, evidenced by higher levels of divergence in exons containing the mature toxin relative to adjacent non-coding regions; this is identical to patterns seen in loci under positive selection in other systems, including snake venom proteins and fertilization proteins in abalone.

(b) extensive gene turnover, with evidence for lineage-specific gene gains and losses in total venom genic content, as well as evidence showing clade-wide extinctions and expansions of specific gene families.

(c) expression regulation, where only 24%–63% of the available venom genes are expressed in adult cone snail species.

Diet and venom evolution: Using extensive ecological data from past studies and comparative phylogenetic methods, I found that while diet specificity (what prey items a species predates upon) does not predict changes in venom gene family size evolution, dietary breadth (how many prey items a species predates upon) is positively correlated with total conotoxin gene diversity. This suggests that species with more generalist diets contain a greater number of conotoxin genes in their genome.

SIGNIFICANCE: First, I successfully applied targeted sequencing techniques to rapidly recover venom gene sequences from genomic DNA at high coverage (> 100X). In a single sequencing run, I generated sequences from > 10,000 venom genes, which is ~2 to 4 times the number of venom genes that have ever been collected over several decades of research. Second, I provided evidence for several genetic characteristics that shape venom composition in cone snails, including positive selection, gene turnover, and expression regulation. These results show that several levels of variation exist and work in concert to ultimately produce the venom phenotype, underscoring the incredible diversity of paths toward generating variation in venom. Finally, I found a positive correlation between dietary breadth and venom gene diversity. These results are consequential because I demonstrate that niche breadth can influence the number of venom genes in a given species' genome, likely having evolutionary consequences for the adaptive potential of species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-381

A complex history of divergence, lineage-specific adaptive evolution, and introgression, accompanies the radiation of Jaltomata species

Meng Wu [1,*], Jamie Kostyun [1], Leonie Moyle [1]

[1]Department of Biology, Indiana University, Bloomington, United States

Abstract: Studies of radiating groups of species can reveal the causes and consequences of diversification, including the genomic substrate of adaptation and the genome-wide signatures of rapid diversification. The plant genus *Jaltomata*—sister clade to *Solanum*—is a morphologically diverse group that has experienced rapid recent evolutionary change in fruit and flower traits, including novel variation in fruit, petal, nectar color, and the size and shape of floral organs. Using whole-transcriptome data from 14 *Jaltomata* species, we perform genome-wide phylogenetic analysis and identify several features associated with radiations. First, although our reconstructed phylogeny shows several well-supported subclades (primarily distinguished by their fruit colors), we find evidence for rampant gene tree discordance due to incomplete lineage sorting (ILS) that arises from rapid successive speciation events. Second, we infer several post-speciation introgression events among the well-supported subclades, which also contribute to observed discordance. Nonetheless, within the context of this phylogenetic complexity, we identify positively selected loci using molecular evolution analyses and, combined with function annotations, generate strong candidate genes for phenotypic diversification and lineage-specific adaptive evolution. Several of these candidates are known players in functionally relevant molecular processes, including pigment biosynthesis loci responsible for fruit color variation.

Expanded summary*: The mechanistic basis of lineage divergence can strongly affect the rate and tempo of diversification. However because the genetic basis of adaptation and speciation is often unknown, the number and nature of loci usually responsible for adaptive differentiation remains debated. For example, in many cases it is unresolved whether rapidly evolving traits are more likely underpinned by a small number of major effect loci or many loci of smaller effect, whether trait differentiation is more often due to mutations in regulatory versus protein coding regions, or whether the birth of new genes plays an important role in the evolution of phenotypic novelty. The recent emergence of next-generation sequencing (NGS) technologies has added a valuable tool to analyses of trait variation. In groups of closely related, ecological diverse species, NGS enables a detailed characterization of clade-wide genetic variation across 100s to millions of genomic sites. In conjunction with information on ecologically-relevant traits, these data can be used to examine the structural, gene expression, and sequence-level molecular variants that might underpin phenotypic diversity and adaptation in these groups. Using a combination of NGS approaches, my dissertation research aims to (i) identify putative genes associated with adaptive traits through phylogenomics, (ii) examine the effect of genome evolution on species-specific phenotypic traits through genome assembly and comparisons; (iii) quantify regulatory evolution and inter-species phenotypic differences using comparative analysis of gene regulatory networks. One goal of my analyses is to dissect the genomic substrate of adaptive evolution in rapidly diversifying clades.

In the presented project, I examine patterns of phylogenetic diversification using clade-wide RNA-seq data in the plant genus *Jaltomata*, a rapidly radiating clade with extensive phenotypic diversity including in floral and fruit traits. My results reveal a complex history of differentiation and gene sharing, due to both incomplete lineage sorting (ILS) and post-speciation introgression. Similar recent findings in other radiating clades suggest these patterns might be a universal feature of rapid adaptation to diverse environmental niches. Within the context of this phylogenetic complexity, I also identify loci associated with lineage-specific trait differences including strong candidate genes for ecologically-relevant traits such as fruit color.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-393

Statistical Package for Growth Rates Made Easy

Portia Mira [1,*], Miriam Barlow [1], Barry Hall [2]

[1]Natural Sciences, UC Merced, Merced, [2]Genetics, Bellingham Research Institute, Washington, United States

Abstract: Growth rates have been an increasingly important method in microbiology because they provide high throughput fitness measurement. The release of Growth Rates Made Easy, a program that uses the output of plate reader files to automatically calculate growth rates, has facilitated experimental procedures in many areas. However, many sources of variation within replicate growth rate data exist and can cause problems with data reliability. We have developed a new statistical package to enhance Growth Rates Made Easy and accurately measure significant variation in growth rate datasets. We found a threshold, Quality-score (Q-scores), which can help determine if variation within a data set is excessive. We have also created a program to creating bootstraps and calculated difference between replicate fitness landscapes. Following these methods will result in dependable growth rate data. These statistical methods are compatible with the analytic methods described in Growth Rates Made Easy and can be used with any set of growth rate output from Growth Rates Made Easy.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS1

Genomic and phenotypic repeatability of divergence between 16 pairs of llake and stream stickleback

Daniel Bolnick*, Yoel Stuart [1]

[1]Department of Integrative Biology, UT Austin, Austin, United States

Abstract: How repeatable is phenotypic evolution? How repeatable is genomic evolution? Do phenotypic and genomic repeatability go hand-in-hand? To answer such questions, we studied 16 replicate pairs of parapatric populations of lake and stream stickleback, from 16 different watersheds. We collected detailed environmental, community, phenotypic, and SNP data from all 16 lakes and 16 streams and 3 marine outgroup populations that represent putative ancestors. Using ddRADseq we obtained >70,000 SNPs from 24 individuals from each of the 35 populations. Phylogenetic and coalescent analyses confirmed the long-suspected independent origin of replicate lake-stream pairs in different watersheds. For each lake-stream pair we identified loci exhibiting exceptional divergence, which may be markers that are linked to genomic targets of divergent selection. To study convergent evolution, we then evaluated how often the same loci are used repeatedly across the 16 replicate lake-stream pairs. Only a few genetic markers were highly divergent in all or most lake-stream pairs. Some between-watershed comparisons exhibited substantial sharing of these supposed targets of selection. Other comparisons between watersheds revealed no significant excess of shared adaptive targets. Notably, watersheds that shared more genomic Fst outliers also tended to have more similar environments, and more parallel lake-stream phenotypic differences. This result demonstrates that genome-wide convergence in Fst-outlier loci coincides with greater phenotypic convergence, as well. And, both genomic and phenotypic convergence are significantly correlated with how parallel the lake-stream environmental contrast is. We infer that convergent genomic evolution occurs but is incompete (for predictable reasons) in stickleback lake-stream pairs.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-390

Detecting polygenic adaptation in an admixture graph

Fernando Racimo [1,*], Jeremy Berg [2], Joseph Pickrell [1]

[1]New York Genome Center, [2]Department of Biological Sciences, Columbia University, New York, United States

Abstract: It is now apparent that much of recent human phenotypic evolution may have occurred via polygenic adaptation (PA): small and concerted shifts in allele frequencies at several loci affecting a complex trait. In recent years, several methods have been developed to detect PA using SNP effect size estimates from GWAS data. Though powerful, these methods suffer from limited interpretability: they can detect which sets of populations have evidence for PA, but are unable to reveal where in the history of multiple populations these processes occurred. To address this, we created a method to detect PA in an admixture graph, which is a representation of the historical divergences and admixture events relating different populations through time. We have developed a MCMC algorithm to obtain posterior distributions of branch-specific parameters reflecting the strength of selection in each branch of a graph. We also developed a set of summary statistics that are fast to compute and can indicate which branches are most likely to have experienced PA. This, in turn, helps us reduce the possible space of candidate branches in our MCMC. We show via simulations that we have good power to detect PA in complex graphs with trait-affecting SNPs from published GWAS data. We also applied our method to human population genomic data from around the world, to determine when and where PA for a variety of anthropometric, metabolic and neurological traits occurred during recent human evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-391

MICROBIALIZER - A SOFTWARE APPLICATION FOR MICROBIAL GENOMICS DATA EXTRACTION AND ANALYSIS

Oren Avram*, Tal Pupko [1]

[1]Cell Research & Immunology, Tel Aviv University, Tel Aviv, Israel

Abstract: The significant technological advances in the last decade brought with them opportunities for large-scale mining and analysis of pathogenic species data in an unprecedented resolution. Such analyses contribute to the comprehensive characterization of complex microbial dynamics within a microbiome and between different strains during a disease outbreak, to name a few. Studying large-scale bacterial evolutionary dynamics poses many challenges. These include data-mining steps, such as gene annotation and orthologs' detection, sequence alignment and accurate phylogenetic tree reconstruction. These steps as well as additional analysis-specific computations require the use of multiple bioinformatics tools and software, making the entire process cumbersome and tedious, and prone to errors due to manual handling.

This motivated us to develop an automatic easy-to-use pipeline that integrates basic and advanced analysis components. In this work, we introduce *MICROBIALIZER*- a user-friendly computational pipeline for the extraction, mining and analysis of large bacterial datasets. Among its features, the user will be able to generate a phylogenetic tree that is based on orthologous set that the program extracts from input genomes. In addition, a module for selective sweeps detection in bacteria (Avram et al., in final preparation) will also be integrated in this pipeline, which should allow scientists to scan for selective sweeps among their input species with a click of a button.

Expanded summary*: Dear committee,

Among all living organisms, bacteria posses an amazing capability for rapid adaptations, impacting various aspects of our lives including health and economy. Thus, unraveling bacterial evolutionary dynamics is of great interest and indeed a large volume of research is devoted to that. Still, large-scale analysis of bacterial evolutionary dynamics poses many challenges. These include data-mining steps, such as gene annotation and orthologs' detection, sequence alignment and accurate phylogenetic tree reconstruction. These steps as well as additional analysis-specific computations require the use of multiple bioinformatics tools and software, making the entire process cumbersome and tedious, and prone to errors due to manual handling.

This motivated us to develop an automatic easy-to-use pipeline that integrates basic and advanced analysis components. In this work, we introduce *MICROBIALIZER*- a user-friendly computational pipeline for the extraction, mining and analysis of large bacterial datasets. Among its features, the user will be able to generate a phylogenetic tree that is based on orthologous set that the program extracts and aligns from input genomes. Of note, this pipeline was needed in my own research, aiming to convince the scientific community that selective sweeps are common in bacteria (Avram et al., in preparation). To the best of our knowledge, this algorithm is the first ever allowing for selective sweep detection in bacteria. We will integrate this algorithm in the proposed software, which should allow scientists to scan for selective sweeps among their input species with a click of a button.

I believe that this work could interest many of the SMBE conference attendance both computational (as some of its components are based on pure computational evolutionary biology) and non-computational (as it facilitates the analysis process of bacterial datasets so that every non-computational biologist should be able to repeat such analyses easily and get insights and conclusions regarding their dataset).

I would be profoundly grateful if the committee were to find me eligible for Young Investigator Travel Award to take part and present my research in the upcoming SMBE conference.

Sincerely,

Oren Avram

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-401

Not only the neo-Y chromosome but also the neo-X chromosome is under accelerated pseudogenization in Drosophila miranda

Masafumi Nozawa [1,*]

[1]Biological Sciences, Tokyo Metropolitan University, Tokyo, Japan

Abstract: Sex chromosomes are derived from a pair of autosomes. After their emergence, recombination between the X and Y chromosomes is generally suppressed, which results in the massive degeneration of the Y through the accumulation pseudogenes and transposable elements. By contrast, little is known about X-chromosome degeneration. This is mostly because the sex chromosomes are ancient in many species, including in model organisms such as humans, mice and *Drosophila melanogaster*, which has hindered reconstruction of the ancestral states of sex chromosomes at their emergence. Consequently, studies of these organisms have not been able to reveal whether the X chromosome degenerates or not. In this study, I therefore utilized the young sex chromosomes, the so-called neo-sex chromosomes in *D. miranda*, and compared the pseudogenization process between genes on the neo-sex chromosomes in *D. miranda* and their autosomal orthologs in a closely related species, *D. pseudoobscura*. The results showed that the pseudogenization rate on the neo-X is much lower than the rate on the neo-Y as expected, but appears to be much higher than the rate on the orthologous autosome in *D. pseudoobscura*. Genes with male-biased expression tend to become pseudogenes on the neo-X, indicating the feminization of the neo-X. Therefore, not only Y but also X may experience accelerated pseudogenization in their early stage of evolution after their emergence. Further studies are necessary to see whether this phenomenon is applicable to a wide range of sex chromosome systems.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-404

Selection on translation rate impacts genic base composition

Erik Quandt [1,*], Charles Traverse [1], Howard Ochman [1]

[1]Integrative Biology, The University of Texas at Austin, Austin, United States

Abstract: The maintenance of a G+C content that is higher than the mutational input to a genome supports the view that selection serves to increase G+C contents in many bacteria. Recent experimental evidence from *Escherichia coli* has demonstrated that selection for increasing G+C content operates at the level of translation, but the precise mechanism by which this occurs is unknown. To determine the substrate of selection, we asked whether selection on G+C content acts across all sites within a gene or was confined to particular nucleotide positions or genic regions. We systematically altered the G+C contents of the GFP gene and assayed the effects of each variant on cellular fitness. The fitness differences were attributable to the base compositional variation in the terminal portion of the gene: increasing G+C content produced more stable mRNA secondary structures, which, in turn, slowed translation rate and allowed proper protein folding. We show that purifying selection against A and T mutations results from their tendency to increase the rate of translation and perturb the dynamics of protein folding.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OM-OS8

The genetic basis of vertebrate pregnancy and birth

Camilla Whittington [1,*], Michael Thompson [2]

[1]Sydney School of Veterinary Science/School of Life and Environmental Sciences, [2]School of Life and Environmental Sciences, The University of Sydney, Sydney, Australia

Abstract: Complex traits such as eyes and live birth (viviparity) are dramatic, adaptive novelties that have shaped the evolutionary trajectories of animals. Viviparity is an important biological innovation that has required a set of complex phenotypic changes to allow internal incubation of embryos, radically changing the way in which organisms interact with their environment and transmit their genes to the next generation. As viviparity has evolved convergently hundreds of times in mammals, reptiles, fish, amphibians, and invertebrates, it is an ideal model to study complex traits, offering the opportunity to compare and contrast naturally replicated evolutionary experiments.

We are using transcriptome profiling coupled with physiological and morphological studies to examine gene expression in the gestational tissues of viviparous animals, and determine the genetic controls underpinning pregnancy in animals that have convergently evolved viviparity. We have identified transcriptional changes associated with nutrient transport, gas exchange, and immunological protection of developing embryos at conception, development and parturition in a range of vertebrates. Key transcripts share homology across pregnant mammals, reptiles, and fish, suggesting a common toolkit of genes regulating pregnancy in divergent evolutionary lineages. Our work shows that common mechanisms may underpin the development of evolutionary innovations across divergent species.

Expanded summary*: My research focuses on using genomics to identify the genes underlying the biology and evolution of novel traits. While the molecular basis of novel traits controlled by a small number of genes, such as antibiotic resistance in bacteria, are well understood, there are significant technical hurdles in understanding the molecular basis of complex novel traits that involve changes in networks of many genes. I apply transcriptomic techniques such as RNA-seq in my work to characterise the genes underlying complex novel traits. In particular, I focus on the genes underlying the genetics of live birth (viviparity), and those genes underpinning venom. My research is significant because it uses genetic techniques to answer broad evolutionary questions (how do complex traits evolve?), whilst also improving our knowledge of the fundamental biology of some of Australia's poorly understood unique native species, including the platypus, pot-bellied seahorse, sharks, and lizards.

Viviparity has evolved more than 150 times in vertebrates, and represents an excellent model system for studying the evolution of complex traits. There are at least 23 independent origins of viviparity in fishes, with syngnathid fishes (seahorses and pipefish) unique in exhibiting male pregnancy. Male seahorses and pipefish have evolved specialized brooding pouches that provide protection, gas exchange, osmoregulation, and limited nutrient provisioning to developing embryos. Pouch structures differ widely across the Syngnathidae, offering an ideal opportunity to study the evolution of reproductive complexity. However, the physiological and genetic changes facilitating male pregnancy are largely unknown. I recently helped to develop the pot-bellied seahorse (*Hippocampus abdominalis*), which has the most complex brood pouch type, as a new research model, by implementing husbandry and culture techniques, constructing an embryonic staging system that is critical to developmental and reproductive research, and developing methods to detect reproductive status (Whittington et al. *J Fish Biol* 2013; Sommer et al. *BMC Dev Biol* 2012). This foundation has enabled our team to use pot-bellied seahorses for transcriptome sequencing *de novo*, generating the first RNA-seq dataset across the full time-course of pregnancy for any animal, and identifying candidate genes underpinning the physiological processes of pregnancy (Whittington et al *MBE* 2015). We examined pouch gene expression at key stages of embryonic development and identified transcriptional changes associated with brood pouch remodeling, nutrient and waste transport, gas exchange, osmoregulation, and immunological protection of developing embryos at conception, development and parturition. Surprisingly, key seahorse transcripts, including *MAPK/ERK* genes and *OVGP1* (implicated in parturition), *MMP19* (implicated in tissue remodelling), and *FABP* genes (nutrient transporters), share homology with genes of reproductive function in pregnant mammals, reptiles, and other live-bearing fish. Syngnathid pregnancy represents the independent origin of a unique form of viviparity in which embryos are brooded by males in a complex brooding organ that is derived from a distinctive tissue type compared to female pregnant vertebrates (abdominal epithelium

versus oviduct).These results therefore suggest a common toolkit of genes regulating pregnancy in divergent evolutionary lineages, which are selected for expression and modified in gestational tissue each time viviparity evolves. We are now using complementary genetic methods to investigate pregnancy genes in other species, which have revealed that while convergence of gene recruitment in pregnancy is important, there may be many genetic pathways to achieve the same viviparous phenotype in amniotes. For example, while some lizards and mammals use the same angiogenic genes to produce uterine blood vessel growth during pregnancy, other closely related squamate lineages use a different set of angiogenic genes (Whittington et al. *J Exp Zool* 2015; Whittington et al. *J Exp Zool* 2017)- showing that convergent gene recruitment is not the whole story of viviparity evolution. In order to explore this area further, I am now broadening our comparisons between independently derived forms of pregnancy to expand well beyond the amniote model, to include RNAseq studies of viviparous sharks (RNAseq data assembly and analysis is currently in progress).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-396

Gene Annotation and identification of microRNA (miRNA) in the salt water crocodile: Crocodylus porosus

Arnab Ghosh [1,*], Roy N. Platt [1], Michael W. Vandewege [1], David A. Ray [1], Sally Isberg [2], Daniel G. Peterson [3], Chuan-Yu Hsu [3], John W. Finger [4], Jaime Gongora [5], Travis Glenn [6], Troy Kieran [6], Rabia tabassum [7]

[1]Biology, Texas Tech University, Lubbock, United States, [2] Centre for Crocodile Research, University of Sydney and Charles Darwin University, Sydney, Australia, [3]Institute for Genomics, Biocomputing & Biotechnology, Mississippi State University, Mississippi State, [4]Department of Biological Sciences, Auburn University, Auburn, United States, [5]Sydney School of Veterinary Science, University of Sydney, Sydney, Australia, [6]Department of Environmental Health Science, University of Georgia, Athens, United States, [7]Faculty of Health, Science and the Environment, Charles Darwin University, Darwin, Northern Territory, Australia

Abstract: Crocodilian genomes have been evolving very slowly over the past several million years, even when compared to their closest extant relatives, the birds. Understanding the evolution, regulation and adaptive capabilities of the crocodilian genome and its genetic diversity can therefore provide information on how slowly evolving genomes manage to stay viable in the face of competition from other taxa and changing environmental conditions. microRNAs (miRNAs) are 21-24-nt sequences that regulate genes via post transcriptional gene silencing and translational repression and are thus important for understanding how gene expression phenotypes evolve. However, no information on crocodilian miRNAs is currently available. Built on a deep sequencing of libraries from 13 unique tissues, utilizing in-house pipelines and established software tools – hundreds of novel miRNAs in the saltwater crocodile were characterized. Many of these originate from transposable elements and show tissue-specific expression. miRNAs in birds and other related species to the saltwater crocodile were also identified. These miRNAs provide novel insights into the evolutionary impact of miRNAs in reptiles as well as regulation of the crocodilian genome.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OM-OS6

Large scale variation in the rate of de novo mutation in humans: quantification, causes and the relationship to divergence and diversity

Adam Eyre-walker [1,*]

[1]School of Life Sciences, University of Sussex, Brighton, United Kingdom

Abstract: The rate of de novo mutation (DNM) in the germline is known to vary across the human genome at a large scale. However, the amount of variation and the causes of this variation remain unresolved. Using a dataset of 45000 DNMs we show that there is significant variation in the rate of DNM across the human genome at both the 100KB and 1MB scales, but that the variation is fairly modest; at the 1MB scale 90% of regions have a rate within 30% of the mean. This variation is similar and correlated for mutations of different types, suggesting that there is coordinated variation in the mutation rate. A multiple regression shows that nucleosome occupancy, recombination rate, replication time and various histone modifications are all independently correlated to the density of DNM, with nucleosome occupancy being the single most important factor. Altogether, the variables explain ~75% of the explainable variance in the density of DNM. The rate of DNM is correlated to both the level of divergence between species and the variation within the human population, but as others have shown for divergence, this correlation is weaker than it could be if all the variation in divergence and diversity was due to variation in the mutation rate. We show that biased gene conversion is partly responsible for this less than perfect correlation, but that it is not the ony factor. We discuss other potential forces that might be at play.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS2

Directed evolution of the genetic code

Andrew Ellington*, Ross Thyer, Drew Tack [1]

[1]NIST, Gaithersburg, United States

Abstract: The genetic code can be expanded through the addition of orthogonal tRNA synthetases and tRNAs that utilize otherwise underused portions of the code, such as stop codons. We have carried out long term evolution experiments that have led to the facile incorporation of two amino acids, selenocysteine and nitrotyrosine, across from amber codons in two different strains of E. coli. In the case of selenocysteine, evolution was done in the context of the so-called 'amberless' E. coli, which has no amber codons in its genome; in the case of nitrotyrosine, evolution was done in a standard lab strain, MG1655. Acceptance of the unnatural amino acids was enforced via engineered 'addiction' to the amino acid in proteins critical for growth. We compare the evolutionary trajectories of these strains and the accommodations that occurred throughout their genomes and proteomes. Improvements in fitness initially accrued through changes in metabolism that likely decreased toxic effects and increased stress responses. The overall fitness of the strains improved, and the new, enforced 21 amino acid genetic codes took hold via the introduction of new amber codons into the genomes of the evolved organisms, although the introduction of essential new amber codons encoding unnatural amino acids was rare. While the canonical code is largely recalcitrant to change, organisms can adapt over time to environmentally induced saltation, as with other long term evolution experiments.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-374

The genetic basis of an ecological transition from freshwater to seawater

Josianne Lachapelle [1,*], Rob Ness [1]

[1]Department of Biology, University of Toronto Mississauga, Mississauga, Canada

Abstract: The adaptive transition between freshwater and marine conditions represents one of the most fundamental ecological transitions as it requires major physiological changes and opens up a diversity of novel opportunities. I have created a unique long-term evolutionary experiment using the freshwater alga *Chlamydomonas reinhardtii* to recapitulate the freshwater to seawater transition. To understand the genetic basis of adaptation to high salt we have sequenced the genomes of the experimental lines that have successfully adapted to marine conditions. In specific we ask (1) What genes are affected and are those genes mutated in independent salt-adapted lines; (2) Do mutations in candidate genes tend to alter protein coding or regulatory sequence; (3) How does the distribution, rate, and spectrum of mutation under selection compare to the patterns expected without selection? We find that each experimental line carries between 4 and 54 unique SNPs, and between 1 and 50 unique indels. The majority of mutations are in genes, with an overrepresentation of mutations in the regulatory UTR sequences. My experimental genomic approach contributes not only in characterizing the genetic basis of an ecological transition but also in providing targets for the engineering of salt tolerance crops and of algae for biofuel production in ocean waters.

Expanded summary*: My research aims to understand the dynamics of adaptation and extinction during environmental change. I combine experimental evolution with molecular and bioinformatics tools to test specific hypotheses and characterise evolution from the genetic to the phenotypic to the fitness levels. My focus has been on the green alga *Chlamydomonas reinhardtii*. Algae play a crucial role in biogeochemical cycles, in the food chain of both freshwater and marine environments, and in the development of biofuels. *C. reinhardtii* is also an excellent system because its genome contains many elements common to land plants and eukaryotes in general, and there are many resources available for metabolic and molecular characterisation.

The adaptive transition between freshwater and marine conditions represents one of the most fundamental ecological transitions as it requires major physiological changes and opens up a diversity of novel opportunities. I have created a unique long-term evolutionary experiment using the freshwater alga *C. reinhardtii* to recapitulate the freshwater to seawater transition. In this experiment, I propagated replicate lines of the alga in increasing concentrations of salt for ~1250 generations. Over 98% of the initial lines went extinct, with only ten lines evolving to survive at salinities exceeding that of typical seawater. I have characterised the fitness of these lines in previous papers. My current research aims to characterise the genetic basis.

To understand the genetic basis of adaptation to high salt we have sequenced the genomes of the experimental lines that have successfully adapted to marine conditions. In specific we ask (1) What genes are affected and are those genes mutated in independent salt-adapted lines; (2) Do mutations in candidate genes tend to alter protein coding or regulatory sequence; (3) How does the distribution, rate, and spectrum of mutation under selection compare to the patterns expected without selection? We find that each experimental line carries between 4 and 54 unique SNPs, and between 1 and 50 unique indels. The majority of mutations are in genes, with an overrepresentation of mutations in the regulatory UTR sequences.

I am currently in the process of identifying the mutations that are shared by salt-adapted lines and comparing the distribution, rate, and spectrum of mutations under selection to the patterns expected without selection. I am also designing experiments to characterize the physiological effects of some of these mutations on salt tolerance. I will cross the salt-adapted lines to a non-adapted ancestor and then compare the metabolites that are excreted by recombinants when grown in high-salt and when grown in freshwater.

My experimental genomic approach contributes not only in characterizing the genetic basis of an ecological transition but also in providing targets for the engineering of salt tolerance crops and of algae for biofuel production in ocean waters.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS7

Ancient genomics of pre-Columbian North American dogs

Laurent Frantz [1,*], James Haile [1], Greger Larson [1]

[1]University Of Oxford, Oxford, United Kingdom

Abstract: Dogs were the first species to be domesticated ~15,000 ago, several thousand years before the advent of settled agriculture. In the New World, the first unequivocal dog remains date to >9,000BP, and previous studies suggested that early American dogs were transported by people arriving from the Old World and were not independently derived from New World wolves. European domestics then largely replaced the early North American dogs, though modern New World dogs may retain a degree of ancestry from the first American dogs. This suggestion remains controversial, however, since the nuclear ancestry of pre-Columbian American dogs has yet to be established. As a result, it remains possible that dogs were domesticated independently in the New World before being replaced by dogs from either Asia or Europe. To test this, we generated four high coverage (and multiple low coverage) genomes from ancient North American dogs including some of the earliest dog remains on the continent (>9,000BP). These data allowed us to assess whether dogs were independently domesticated in North America, when they were replaced by Eurasian dogs, how many times they were replaced, and the degree of extant pre-Columbian ancestry in modern North American dogs. Lastly, this data allow us to characterise the demographic and adaptive history of this isolated population of domestic dogs.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-408

Temporal dynamics of mutations that affect GC content in bacteria

Saurabh Mahajan [1,*], Deepa Agashe [1]

[1]National Centre for Biological Sciences, Bangalore, India

Abstract: The proportion of GC basepairs in bacterial genomes varies from 15-75%. It is now established that mutations are inherently biased towards AT, and high GC content arises from a process favoring GC. However, many issues about the diversity and the underlying processes remain unresolved. Does the extent of the inherent mutational bias differ across bacteria and does this contribute to differences in GC content? On what time-scale does the directional process act on GC? We investigated the temporal dynamics of mutations affecting GC content in multiple sets of closely related bacteria. For every set, we first counted changes from GC→AT and AT→GC on each branch of a phylogeny. On short branches, the mutational bias must appear as excess of GC→AT over AT→GC changes. On longer branches, the excess must erode as process(es) favoring GC have time to act, leading to a balance between changes in both directions. We analyzed excess GC→AT changes as a function of branch length to estimate the mutational bias and the time taken for a balance to be reached. Across multiple bacteria, we find that the mutational bias is correlated with existing GC content. Surprisingly, GC basepairs are favored even in some low GC bacteria. Second, excess GC→AT changes are lost as lineages diverge by 1-10% of their genome sequence. We conclude that inherent mutational biases do contribute to GC content diversity, and that any selective forces favoring GC can act on small differences of <1% in GC content.

Expanded summary*: That bacterial genomes differ in their GC content (15-75%) has been recognized for over 50 years. Yet, how this diversity is generated is not exactly understood. Seven years ago, it was convincingly established that mutations generate excess GC→AT changes in almost all bacteria. These findings implied that high GC content must evolve under a directional force that favors GC. However, this alone does not resolve the issue of what generates differences in GC content. Differences could arise because- mutational biases differ and selection is constant, or selection differs and mutational biases are constant, or both differ. This issue has remained unresolved (Rocha and Feil; PLoS Genetics 2010). We provide more support for the idea that differences in mutational biases contribute to differences in GC content. In a dataset containing multiple bacteria, we show that mutational bias is correlated with existing GC content. Surprisingly, we also found that GC basepairs are favored even in some low GC bacteria. This contradicts the current understanding that low GC arises from the weakening of selection for GC (like in endosymbionts). Instead, we find that in some cases, the low GC can be explained by a stronger mutational bias towards AT. This deepens the mystery of what might generate selection on GC content and why mutational biases differ across bacteria. Finally, by estimating the speed of the process favoring GC, we suggest that selection may be able to act on differences smaller than 1% in GC content. To our knowledge, this is the first time that such limit has been identified. This raises the following question- which phenotypes relevant for selection may vary appreciably with <1% change in GC content? Overall, our study provides support for a hypothesis explaining GC content diversity, identifies counter-examples to the current understanding, and raises new questions.

In our study, we have adopted an approach based on previous studies (Balbi, Rocha, Feil, MBE 2009; Rocha et al, 2006 J Theor Biol). This approach makes use of the signatures of mutation and selection implicit in the temporal data on mutations. We began by inferring genome-wide nucleotide changes occurring on the phylogenies of multiple sets of closely related bacteria. For this purpose, we implemented an averaging weighted by posterior probabilities (AWP) approach to reconstruct ancestral changes. Then, we quantified the excess GC→AT changes as a function of branch lengths. We expect that on short branches, excess GC→AT changes reflect the inherent mutational bias towards AT; while on long branches, the directional process gradually purges excess GC→AT changes. Quantifying the relation between excess GC→AT changes and branch length allowed us to obtain estimates of the inherent mutational bias, and importantly, to identify the speed of the directional process favoring GC. To our knowledge, this is the first time that a rigorous genome-wide ancestral reconstruction method and subsequent temporal analyses were used to study the evolution of GC content. We believe that this approach can also be used to study the related problem of the evolution of AT skew (differential usage of A versus T).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-370

Recent Himalayan population demographic history: insight from whole genome deep sequencing data

Elena Arciero [1,*], Thirsa Kraaijenbrink [2], Asan . [3], Marc Haber [1], Qasim Ayub [1], Mark Jobling [4], George Van Driem [5], Yali Xue [1], Peter De Knijff [2], Chris Tyler-Smith [1]

[1]Wellcome Trust Sanger Institute, Hinxton, United Kingdom, [2]MGC Department of Human and Clinical Genetics, Leiden University Medical Centre, Leiden, Netherlands, [3]BGI-Shenzhen, Shenzhen, China, [4]Department of Genetics, University of Leicester, Leicester, United Kingdom, [5]Himalayan Languages Project, Institut für Sprachwissenschaft, University of Bern, Bern, Switzerland

Abstract: Himalayan populations reside at a broad range of altitudes and in different environments, ranging from tropical forest to high mountain peaks. Isolation, genetic drift and natural selection have shaped the genetics of these populations. We are using a combination of SNP-chip data and genome sequences to explore these influences. We used ~600,000 genome-wide SNPs typed in 948 Himalayan individuals from 49 different autochthonous groups from Nepal, Bhutan, North India and the Tibetan Plateau in China to show that Himalayan populations share a component derived from a common ancestral population, followed by the development of local fine genetic structure correlating with language and geographical distribution. We used genome sequences of 88 individuals from a subset of the Himalayan populations residing at different altitudes and environments to refine our current knowledge of their demographic history, using both autosomal and uniparental markers. We also explored how isolation and genetic drift might have influenced the frequencies of rare (and probably recent-origin) variants in Himalayan populations and we used these rare variants to understand more recent history. Genetic signatures of adaptation to high altitude in *EPAS1,* and signatures of adaptation to low altitude within the *TRIM67* region, possibly associated with anti-microbial activity in the tropical forest, have been detected using the genome-wide SNP data. We are using the genome sequences to refine these selection signals and narrow down possible functional candidate variants for further validation. We are also looking for possible new signatures of adaptation in Himalayan populations not previously included in our SNP-chip dataset.

Expanded summary*: Himalayan populations have settled in a broad range of altitudes and environments that have been played a strategic role in shaping their genetic, cultural and ethno-linguistic mosaic. We generated one of the biggest up-to-date dataset of ~600,000 genome-wide SNPs typed in 948 Himalayan individuals from 49 different autochthonous groups from Nepal, Bhutan, North India and the Tibetan Plateau in China and genome sequences of 88 individuals from a subset of the Himalayan populations including three populations not present in our genome-wide SNPs data. We were able to genetically characterise populations from the Himalayan region that have not been extensively studied before, disclosing their population structure and demographic history. We also found genetic signatures of adaptation to high altitude in *EPAS1,* and signatures of adaptation to low altitude within the *TRIM67* region, possibly associated with anti-microbial activity. The genome sequencing data gave us a better understanding of more recent demographic history and the possibility to explore how isolation and genetic drift could affect the frequency of rare variants in these populations. Moreover, we could refine the selection signal detected using SNP-chip data at both high and low altitude and look for other signals of adaptation in the new populations available to narrow possible functional candidate variants for validation. Although many publications reported *EPAS1* as under positive selection for high altitude adaptation, functional study on *EPAS1* variants has not systematically carried out and it is still unknown which are the variants responsible of high altitude adaptation and their mechanism of action. We use both *in silico* and *in vitro* studies to explore the molecular mechanism of high-altitude adaptation in Himalayans and validate *EPAS1* candidate regulatory variants. We are also investigating whether Himalayan populations show more efficient system of DNA repair due to higher volume of ultraviolet (UV) radiation and hypoxia with the production of reactive oxygen species (ROS) at high altitude. This research combine computational and laboratory work to describe human genetic variation in populations not extensively represented at the worldwide level and to unravel molecular mechanism of human adaptation in different environments. Furthermore, this dataset could become a very useful resource for other researchers interested in population genomics and disease association studies.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-369

Evolution of the sex determining systems in the genus Silene - section Otites

Roman Gogela*, Veronika Balounova, Jitka Zluvova, Jan Safar, Bengt Oxelman, Radim Cegan, Roberta Bergero, Deborah Charlesworth, Roman Hobza, Boris Vyskot, Bohuslav Janousek

Abstract: Switches in heterogamety occasionally occur both in animals and plants. In contrast to animals, sex determination in plants is usually of more recent origin and so there has been shorter time-frame for the putative switches. Our previous research revealed a switch in heterogamety from XY to ZW or *vice versa* in the section Otites in the plant genus Silene.
In this study, we have performed a detailed analysis of the evolution of sex determining systems in the section Otites including the analysis of the ancestral status of heterogamety in this section. Further we have tested if the sex chromosomes of species with different type of heterogamety evolved from different pairs of autosomes.
Our current data suggest that female heterogamety was the original sex determining system in the section Otites. Female heterogamety was confirmed in *S. otites* and newly revealed in *S. borysthenica*. Results of the genetic analysis and mapping of the sex-linked sequences showed that sex chromosomes of the species showing female heterogamety (*S. otites* and *S. borysthenica*) evolved from different pair of autosomes than sex chromosomes of *S. colpophylla* that shows male heterogamety. The change of female heterogamety to male heterogamety in the section Otites represents the first such case described in plants.
This research was funded by Czech science foundation (No. 1700567S)

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OM-OS2

The number of independent organelle DNA insertions in genomes

Einat Hazkani-Covo [1,*], William F Martin [2]

[1]Department of Natural and Life Sciences, The Open University of Israel, Raanana, Israel, [2]Institute of Molecular Evolution, Heinrich-Heine University, Düsseldorf, Germany

Abstract: Fragments of organelle genomes are often found as insertions in nuclear DNA. These fragments of mitochondrial DNA (numts) and plastid DNA (nupts) are ubiquitous components of eukaryotic genomes. It is now recognized that they occur during cancer progression as well. Numts and nupts, once inserted, can become further fragmented through subsequent insertion of mobile elements or other recombinational events that disrupt the continuity of the inserted sequence relative to the genuine organelle DNA copy. Because numts and nupts are typically identified through sequence comparison tool such as blast, disruption of insertions into smaller fragment can lead to systematic overestimation of numt and nupt frequencies. Accurate identification of numts and nupts is, however, important, not only for better understanding their role during evolution, but also for assessing their increasingly evident role in human disease. Here we report investigation of salient parameters involved in obtaining accurate estimates of numt and nupt numbers in genome sequence data. Numts and nupts from 47 sequenced genomes reveal lineage specific differences in the number, relative age and frequency of postinsertional fragmentation of insertional events and circumscribe the main parameters influencing accurate identification and frequency estimation of numts in studies pertinent to human health.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-373

Thermal adaptation in an emerging hybrid species

Elzbieta Iwaszkiewicz*, Arne W. Nolte

Abstract: Hybridization can lead to the evolution of new species but many biological details underlying such processes are unclear. We study a recently emerged hybrid lineage of *Cottus* fish, which has invaded big river habitats downstream of the areas inhabited by their parental species. The ecological differences between the relevant upstream and downstream habitats involve vastly different animal and plant communities, but are ultimately attributed to different light and temperature regimes. Here, we explore whether temperature differences are likely to play a decisive role in the evolution of invasive *Cottus*. We used RNA-Seq to compare parental species and the hybrid lineage in controlled laboratory conditions. This analysis was extended to an unprecedented depth by analyzing time series for 96 fin and 96 liver transcriptomes from wild fish, sampled in the course of a complete year. We find that gene expression divergence is most pronounced during the summer and identified clusters of genes that respond differentially to temperature between parent and hybrid. The prevalence of temperature dependent patterns and divergence suggests that thermal adaptation, especially during summer conditions, is a key force in the evolution of invasive *Cottus*.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-325

The effects of sex-biased gene expression and X-linkage on rates of adaptive and non-adaptive sequence evolution in Drosophila melanogaster

José L Campos*, Keira Johnston, Brian Charlesworth [1]

[1]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

Abstract:

A faster rate of adaptive evolution of X-linked genes compared with autosomal genes (the *Faster-X effect*) may be caused by the fixation of recessive or partially recessive advantageous mutations. This effect is expected to be largest for advantageous mutations that affect only male fitness, and absent for mutations that affect only female fitness. We tested these predictions in *Drosophila melanogaster* by using coding and non-coding sequences of genes with different levels of sex-biased expression, estimating the extent of adaptive evolution of nonsynonymous and non-coding mutations from polymorphism and divergence data.

Consistent with the theoretical expectations, nonsynonymous substitutions in coding sequences of most male-biased and unbiased genes evolve faster on the X. However, genes with very low recombination rates do not show a Faster-X effect, possibly as a consequence of Hill-Robertson interference. Contrary to expectation, there was a Faster-X effect for strongly female-biased genes. However, these genes had higher recombination rates than comparable autosomal genes, and higher recombination rates are known to be associated with faster rates of adaptive evolution. After correcting for recombination rate, strongly female-biased genes do not show a Faster X-effect. Similar analyses of non-coding UTR and long intron regions showed a Faster-X effect for all groups of genes considered. The Faster-X effect for non-coding regions is likely to be due to cis-regulatory mutations that affect expression levels.

In contrast, given the strong evidence that deleterious mutations are mostly recessive or partially recessive, we would expect a slower rate of evolution of X-linked genes for mutations that are non-adaptive and fixed by genetic drift. Surprisingly, we found little difference between the X and the autosomes in the rates of fixation of non-adaptive mutations. We discuss possible reasons for this.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OM-OS4

Tracking bacterial genetic variation over short time spans with metagenomics

Falk Hildebrand [1,*], Toni Gossmann [2], Jaime Huerta-Cepas [1], Sebastian Waszak [1], Anna Pryszlak [1], Ana Zhu [1], Sonja Blasche [1], Peer Bork [1]

[1]EMBL, Heidelberg, Germany, [2]University of Sheffield, Sheffield, United Kingdom

Abstract: Recent advances in metagenomics have revealed considerable genetic variation among the microbes that populate the human gut. While it has been shown that multiple strains of the same species can coexist in the same microenvironment, it is still under debate how stable this is.
In this work we were able to reconstruct the genomic sequence of a highly abundant yet previously uncharacterized bacterial species from the human gut microbiota, that coincided with this patient being treated with antibiotics. Using this workflow, we additionally de-novo assembled genomes of selected highly-abundant bacteria. For these, single nucleotide variations (i.e. SNVs) spanning multiple time points were determined. After a strict SNV quality filtering procedure that removed >99% of SNV calls, we describe the evolution of bi-allelic loci and their propagation and/or loss over 3 years time within each bacterial species. Using allele frequency trajectories we can show that the effective population size varies considerably between and within bacterial species, with some of these stably colonizing the gut for extended periods.

Expanded summary*: In microbial genomics, classical reductionist approaches, focusing on single species, genes and genomes, are increasingly replaced by holistic "whole-community" metagenomic studies. The metagenomics field has gained in importance over the last years, mainly due to a) new sequencing technologies enabling the field and b) numerous associations between microbial organisms and their environment being made. For example, the diversity of fungal diversity and associations to abiotic parameters can be described using metagenomics, but on the other spectrum complex human diseases like Inflammatory bowels disease can be associated to a gut microbial dysbiosis. Thus, this field and its discoveries has the scale to have a direct impact on human health, as well as describing vital parts of the world ecosystem.
In the metagenomic approach, DNA of all genomes found in a given ecosystem is randomly sequenced, and the genomic content of microorganisms within the environment are reconstructed, in order to enable quantification of taxonomical and functional diversity within a given community. Determining the exact strain composition within a community is not straightforward, due to sequencing errors and theoretical difficulties in defining strains, just based on short read sequences randomly sampled from a given environment. Recent advances in bioinformatic algorithms, using single nucleotide variants (SNVs), allow for the distinction of single bacterial species, or even strains. These advances are critically important, as pathogenicity and metabolism of bacteria are often strain specific. SNVs can also guide the reconstruction of the bacterial genomes, thus linking the gained increased in taxonomy to strain-specific functions.
In this work we exploit single nucleotide variants (SNVs) in the human gut microbiome to identify and delineate distinct bacterial strains. We were able to reconstruct the genome of a highly abundant yet previously uncharacterized bacterial species found in an antibiotics-treated patient. For this genome and additional de-novo assembled genomes of known species, we estimate SNVs over time, monitoring strain variability and nucleotide variations over time. Further we were also interested in comparing the same species, but different dominant strains, across the microbiome of several patients within our dataset.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-187

Population differentiation and molecular evolution of pheromone communication in two orchid bee sibling species

Philipp Brand [1,*], Thomas Eltz [2], Santiago Ramirez [1]

[1]Department of Evolution and Ecology, UC Davis, Davis, United States, [2]Evolutionary Ecology and Animal Biodiversity, Ruhr-University , Bochum, Germany

Abstract: Pheromone communication has been long known to play a central role in the origin and evolution of species diversity throughout the tree of life. However, the underlying genetic and molecular mechanisms that control pheromone variation and signal detection remain poorly understood. We investigated the genetic basis of signal production and detection in orchid bees. Unlike the majority of insects, male orchid bees collect chemical substances from various environmental sources to concoct complex pheromone mixtures that are subsequently exposed to females during courtship display. Male-gathered pheromones are species-specific and highly divergent among taxa, suggesting they play a key role in the origin and maintenance of reproductive isolation among lineages. We investigated the genetic mechanisms underlying changes in pheromone composition and chemosensory detection, and the role they play in generating and maintaining reproductive isolation in two recently diverged orchid bee sibling species. Therefore, we conducted a population analysis including population genomic, pheromone chemotype, and chemosensory receptor resequencing analyses based on >200 bees of 17 sympatric and allopatric populations throughout the distribution ranges of both species. Our results indicate that several chemosensory receptor genes are highly differentiated among species and exhibit strong signatures of divergent selection, a common pattern known for genes important in the process of speciation. Further, genetic differentiation of one receptor correlates with sympatric - allopatric species boundaries, congruent with the hypothesis of molecular character displacement on the gene. Our results support the hypothesis that chemosensory receptor gene divergence could underlie the evolution of reproductive isolation in orchid bees.

Expanded summary*: From bacteria to mammals, living organisms of all levels of complexity use chemical signals to detect, discriminate and choose mates. Ever since the first sex pheromone was characterized in the silkmoth in 1957, a goal of evolutionary biology has been to explain how pheromone communication systems diverge and how their evolution contributes to species diversity. However, despite the ubiquity and importance of pheromones in animal communication, the underlying genetic and molecular mechanisms that control signal chemistry and detection remain poorly understood.

My dissertation research focuses on the study of the insect pheromone system of orchid bees and aims to elucidate how natural selection shaped the evolution of signal production and signal detection.

Orchid bees are among the most important pollinators of thousands of diverse plant species in tropical America. Unlike most other insects, male orchid bees do not produce their own pheromones internally but instead detect and collect chemical substances from various floral and non- floral sources to concoct species-specific pheromone blends, which are uniquely attractive to female bees of the same species. As a result, orchid bees rely exclusively on olfaction to produce and detect species-specific chemical signals. It is therefore likely that chemosensory genes simultaneously important for pheromone acquisition and recognition are subject to novel evolutionary pressures when signals change.

For my dissertation research I am investigating the genetic mechanisms of pheromone divergence and the role that chemosensory detection and olfactory processing play in the evolution of reproductive isolation. At SMBE 2017 I plan to present my work on testing the hypothesis that chemosensory receptor (*CR*) divergence is responsible for phenotypic pheromone variation and reproductive isolation between orchid bee lineages.

To test my hypothesis, I conducted a population analysis of the two recently diverged orchid bee sibling species *E. dilemma* and *E. viridissima* (diverged *ca.* 0.17 Mya) distributed throughout Central America with overlapping sympatric distributions for about half of the respective distribution ranges. I sampled bees at 17 allopatric and sympatric sites throughout the respective distribution ranges from southwestern Mexico to northern Costa Rica. Collected individuals were used in a population genomic, pheromone chemotype, and targeted *CR* resequencing analyses. For the population genomic study, I sequenced neutral SNPs throughout the genome in a total of 285 individuals using Genotyping-by-Sequencing. The analysis of the resulting 37,679 SNPs revealed low divergence between populations and species, reflecting high connectivity between populations and the low divergence time between the two species.

Despite this similarity at neutral genetic markers, individuals were clearly distinct in their pheromone chemotypes and could be unambiguously classified into the respective species. Next, I sequenced *CR* genes in the same individuals, which I previously identified in a comparison of the species' antennal transcriptomes enriched for CR mRNA. This way, I could show that several *CR* genes are highly differentiated among species and exhibit strong signatures of divergent selection, a common pattern known for genes important in the process of speciation. Further, genetic differentiation of one receptor correlates with sympatric - allopatric species boundaries, congruent with the hypothesis of molecular character displacement on the gene. Overall, these results suggest a role of *CR* divergence in the evolution of reproductive isolation between *E. dilemma* and *E. viridissima*.

I am currently working on the functional characterization for the CR genes under strong diversifying selection between the two species. Using the *Drosophila* empty-neuron system, I am expressing target orchid bee receptors in olfactory sensilla in the sensory system of transgenic flies, which I test using neurophysiological single-sensillum recordings. I will include potential results in my SMBE talk if possible.

My dissertation research contributes to an only recently growing field of evolutionary biology. The molecular evolution of pheromone communication systems is an underexplored topic integral to understand the origin and maintenance of species and thus biodiversity on earth. By integrating molecular genomic, chemical and functional neurophysiological analyses in a population biological framework, I am applying a novel approach to shed light on the long asked question of how pheromone communication systems evolve. My dissertation work will advance our understanding of molecular mechanisms, emphasizing the evolution of pheromone communication, and provide insights into the process of speciation. Moreover, this work extend the field of molecular pheromone research to a formerly neglected, yet, very important group of insect pollinators, the bees. Overall, the results of my work will highlight the significance of combining population biological approaches and pheromone research for studies on non-model organisms. This research has the potential to also illustrate the function of key genes involved in reproductive isolation in orchid bees, leading to a deeper understanding of the great diversity in this group of important plant pollinators.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-386

Methodological aspects of microbiome analysis of insects as pathogen vectors

Sonia M. Rodríguez-Ruano [1,*], Vaclav Hypsa [1 2], Eva Novakova [1 2]

[1]Department of Parasitology, University of South Bohemia, [2]Institute of Parasitology, Academy of Sciences, Ceske Budejovice, Czech Republic

Abstract: The microbiome may affect competence for pathogen transmission in insect vectors. However, the knowledge on the vector-microbiome-pathogen interactions remains limited. For instance, current methodologies for mosquito surveillance provide reliable information about infection prevalence and distribution, but fail to correlate specific mosquito holobiont traits (i.e. microbiome) and infection status due to sample pooling. On the other hand, studies focusing on mosquito microbiomes have so far been limited to small sample sizes due to the cost of molecular procedures and sequencing. When natural (instead of laboratory-reared) populations are used, the sample size rarely suffices to retrieve epidemiological data due to the generally low prevalence of arboviruses vectored by mosquitoes. Thus, the studies of mosquitoes as viral vectors and as bacterial hosts are to some extent uncoupled. This study focuses on finding a suitable combination of methodologies to reduce the cost of extensive sampling while allowing for individual correlations between microbiome and infection status in these important pathogen vectors. Our sampling included *Aedes vexans* mosquitoes captured in South Bohemia (Czech Republic) and southern Ontario (Canada) between 2011 and 2016. The methodologies tested comprise three key aspects that affect cost and feasibility of broad scale molecular studies: pooling, sample preservation, and dissection of specimens. Czech mosquitoes were individually preserved either in AllProtect (Qiagen), a lab-made solution (NAP), or ethanol, and stored at 4°C or -20°C for one week prior to DNA/RNA extraction. Fresh samples were used as a control for the preservation effect, and also for dissection; the tested protocols included DNA/RNA extraction from whole mosquitoes, guts, and the rest of the dissected body. The effect of pooling was tested with different pool sizes (1-50 individuals) using mosquitoes from Canada. Preservation treatment and dissection did not significantly affect the microbiome diversity found, while pooling posed a major effect. Our recommended protocol for studies on vector-microbiome-pathogen interactions includes individual mosquito preservation in NAP solution at 4°C, whole specimen DNA/RNA extraction and subsequent RNA pooling for initial screening of viruses prior to high-throughput sequencing of the microbiomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-339

Natural social adaptation in a social amoeba revealed by signatures of molecular evolution

Suegene Noh [1,*], Katherine Geist [1], Joan Strassmann [1], David Queller [1]

[1]Washington University in St. Louis, Saint Louis, United States

Abstract: Microbes have come to play an important role in the study of how cooperation and conflict drive social behaviors, largely because they can be genetic manipulated and experimentally evolved relatively easily. Whereas animal social behavior can be observed and assessed in their natural environments, this is not the case for microbes. As a result, we know little about microbial social adaptations in nature. This has led to some difficult-to-resolve controversies about social adaptation even in well-studied microbial phenomena such as bacterial quorum sensing and siderophore production, and social amoeba altruism and cheating. Here we illustrate one way to address this problem using signatures of molecular evolution. Specifically, we focus on the impact of cooperation and conflict during multicellular transition in social amoebae (*Dictyostelium discoideum*), during which tens of thousands of individual amoebae aggregate and form a multicellular fruiting body. We show evidence of positive selection driving evolutionary arms races in genes that are up-regulated in the presence of foreign clones. This indicates that social cheating has a significant evolutionary impact in nature. We also used molecular evolutionary signatures of conditional selection and kin selection to show that pre-stalk cells, that create the structural element of the fruiting body but cannot reproduce like pre-spore cells, evolve through altruistic kin effects rather than solely by selfish direct effects. This emphasizes the role of kin selection in nature.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-342

Effective population size drives the evolution of gene repertoires of Bacteria

Louis-Marie Bobay [1,*], Howard Ochman [1]

[1]Integrative Biology, University Of Texas, Austin, United States

Abstract: Effective population size (*Ne*) is a key parameter in population genetics that dictates the relative impact of drift and selection acting on natural populations. Species undergoing reduced *Ne* experience a relaxed effectiveness of selection, which diminishes their ability to counteract the accumulation of detrimental variants. Estimating *Ne* is a sensitive endeavor in bacterial species due to i) the difficulty of identifying truly neutral sites, ii) the loss of diversity mediated by the strong linkage acting along bacterial genomes, iii) the arbitrariness of species definitions, and iv) the unevenness of divergence rates across species. In this study we computed direct and indirect estimates of the effective size of 153 prokaryote species under a unifying framework. Our results indicate that bacterial ecology is a major component driving *Ne*, but, in contrast to theoretical predictions, recombination rates were not found associated with variations of *Ne* estimates. Finally, our analyses revealed that *Ne* drives the evolution and diversity of gene repertoires in bacteria—not genome sizes. Altogether, these results point toward the edification of a new model of the evolution of genome architecture in bacteria.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-340

Defining the mechanisms by which miR-506-3p regulates MYCN expression

Spencer Shelton [1,*], Liqin Du [1], Zhenze Zhao [1]

[1]Biochemistry, Texas State University, San Marcos, United States

Abstract: Neuroblastoma, an aggressive childhood cancer, can be diagnosed by over expression of known oncogenes. MYCN, which is over expressed in 20% of high risk neuroblastoma cases, is a genetic alteration that plays a key role in tumorigenesis through regulation of cellular differentiation. We previously identified miR-506-3p as a strong repressor of MYCN as well as an inducer of neuroblastoma cell differentiation. The 3'UTR of MYCN mRNA lacks the target site of miR-506-3p, suggesting that miR-506-3p regulates MYCN expression through an indirect pathway. By combining gene expression array and informatics analyses, we identified 11 candidate genes in the miR-506-3p pathway that target MYCN. We performed qPCR to investigate the possible role of each gene in regulating MYCN expression. We performed a high content screen on siRNAs to investigate the effect of each gene knockdown on neuroblastoma cell neurite outgrowth - the morphological differentiation of neuroblastoma cells. Interestingly, only one gene, RXRA, showed significant down regulation of MYCN when knocked down. Furthermore, knockdown of RXRA showed the most dramatic effect in inducing cell differentiation among the 11 genes. Further investigation is needed to examine the pathway and molecular mechanisms through which RXRA mediates neuroblastoma differentiation and MYCN expression.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-330

Evolutionary Probability is robust to breadth of taxonomic sampling

Ravi Patel*, Sudhir Kumar [1]

[1]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, United States

Abstract: Evolutionary Probability (EP) captures the neutral evolutionary expectation for the presence of an amino acid residue (or nucleotide) at a given position in a genome. It uses interspecific evolutionary sequence alignment to make a null expectation prediction, which can then be used to determine non-neutral (disease or adaptive) in a population. Proper application of the Bayesian method used to estimate EP requires that a large number of diverse sequences be used. But, no information exists on the taxonomic breadth and sampling that will produce robust EP estimates. Here we report results on the evaluation of the robustness of EP estimates from temporal and density sampling of species, with a focus on the human genome. We find that the temporal depth of the species included in the alignment is more important than the larger density of species sampled closer to humans (e.g., apes or primates). We show that the examination of the patterns of EP estimates under temporal sampling at fast evolving positions in a protein provides a way to detect if the sequence alignment contains enough data to estimate EP robustly. We will also present results from analysis of data from microbial and bacterial proteins, which makes our contribution widely applicable.

Expanded summary*: Recently, the evolutionary probability (EP) method was proposed as a way to generate expected probabilities for the observation of an allele at a given position in a protein in a species. The EP value is calculated by using a Bayesian approach for inferring the likelihood of the presence of a residue given the long term evolutionary history of a species, and ranges from 0 to 1. In an analysis of the human proteome, Liu et al. (2016, Mol Biol Evol 33 (1): 245-254) showed that EP correlates well with population allele frequencies. They proposed that EP can be used as a null expectation for an allele's frequency in a population. This relationship can thus be used to detect variation with strong negative effects, since an allele that is unexpected under the EP framework likely has a relatively negative consequence on fitness (compared to the alternate or more likely allele).

The EP method can be applied to any species for which evolutionary history is available to find such variation. Even for species with little or no polymorphism data, the method can be used to create a catalog of potentially damaging variation and as a reliable mutation diagnosis method. Alleles with high EP should be found at high allele frequency, while alleles with low EP should be found at low frequency. Deviations from this trend can be indicative of potentially non-neutral (e.g., adaptive) changes in a species as well. Of course, non selective forces such as demography as well as the timing of origin for a variant can play a role in creating discrepancies between EP and allele frequency, however most population genetic based analysis are subject to this issue.

While the original method developed by Liu et al. was created using a very large collection of divergent species (46 vertebrates,> 600 million years evolutionary divergence), the method can be applied to any set of taxa with sufficient evolutionary information. Also, EP does not require population level information, and thus can provide information for all sites regardless of their polymorphism state in a population; e.g., evolutionary probability can also be calculated for unobserved residues at fixed and evolutionarily conserved positions. However, contrasting the EP of a residue with its allele frequency (AF) can presumably inform on its current state of selection, e.g., low EP alleles at high AF are potentially adaptive in a population (cite).

No test of the method was conducted on the adequacy of the dataset given the expansiveness and density of data available for human and related species protein sequences. For the generalization of this method, we must have a way to evaluate the reliability of calculated EP values for a given dataset for more robust analyses and prioritization of function-impacting variation in a species. Further, we must know about the biases introduced by inadequate data sampling. Thus, the determining and setting of guidelines for proper EP calculation using sufficient data is the focus of this study, as well as determining the robustness of EP calculation. We investigated the effects of sequence compositions, alignment quality, and evolutionary time spans for a dataset to determine the minimal dataset required for application of the EP method.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-328

Transposable elements and lineage sorting within the genus Myotis

Jennifer Korstian [1,*], Neal Platt [1], David Ray [1]

[1]Biological Sciences, Texas Tech University, Lubbock, United States

Abstract: The exact evolutionary relationships within the genus *Myotis* remain unresolved because phylogenies constructed with nuclear and mitochondrial genes produce different topologies. A number of factors can drive conflict between markers, such as hybridization, phylogenetic error, and incomplete lineage sorting. Transposable elements (TEs) are DNA sequences that have the ability to move around in the genome and are powerful genomic markers. We examined the genomes of 9 North American bats from the genus *Myotis* and used the Mobile Element Locator Tool (MELT) to identify differential insertion patterns of five TE families in each species. We've mapped those insertions onto the most recent nuclear phylogeny of the clade. Preliminary results suggest that the early, rapid diversification was associated with significant lineage sorting.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-324

Genomic Reconstruction of The Last Eukaryotic Common Ancestor.

David Newman [1,*], James McInerney [1]

[1]Division of Evolution and Genomic Sciences, University of Manchester, Manchester, United Kingdom

Abstract: The current breadth of available eukaryotic genomic data provides us with the opportunity to better understand the origin and evolution of the eukaryote. Contemporary eukaryote genomes contain a mixture of genes: some of whom can be traced back to prokaryotic ancestors; some whose origins are likely to be coincidental with the origin of the eukaryotic cell; and some that have arisen more recently. Differential losses of genes have balanced gene gains to a certain extent. However, this complex process makes it difficult to unravel the history of eukaryotic genomes and to reconstruct the genome of the last eukaryote common ancestor (LECA). We have identified and analysed widely conserved gene families with broad distributions in eukaryotes and carried out metabolic reconstructions in order to provide a clearer picture of eukaryote origins and evolution.

Expanded summary*: The project seeks to assemble an, as complete as physically possible, list of genes likely to have been present in the Last Eukaryote Common Ancestor (LECA). This is being achieved through utilizing the current wealth of complete genome data for eukaryotes to find conserved gene families and performing clustering and phylogenetic analyses upon them to tease apart signals of vertical as opposed to horizontal inheritance. The evolution and origins of eukaryotes is an active and broadly appealing field of research. The recent identification of archaea, the Asgard Archaea, which represent the prokaryotes that are most closely-related to eukaryotes being published in Nature as an example.

In order to reconstruct the genome of the last eukaryote common ancestor, a list of 32 genomes from publically available data was acquired. This list was constructed to represent the six major clades amongst the eukaryote kingdoms of life, namely: the Amoebozoa; the Excavata; the Archaeplastida; the Orphans; the Opisthokonts; and SAR, as well as reflecting as much of the phylogenetic diversity within each of these said groups as possible. A list of over 100 bacterial and archaeal genomes, or partial genomes, if a species of sufficient interest, such as the aforementioned Asgardian Archaea, was also generated representing a range of the prokaryotic taxonomic diversity for comparison against the eukaryotes.

Homologous gene families from the combined list of eukaryotic and prokaryotic genomes were identified with Blast and the output was clustered using markov cluster Algorithm (mcl) analysis. These clusters were then filtered, first, by the presence of genes from three or more of the eukaryotic clades, then secondly categorizing the clusters that show sufficient taxonomic diversity as belonging to either, (i) Eukaryote-Specific Proteins (ESPs), (ii) Eukaryote/Eubacterial homologs, (iii) Eukaryote/Archaebacterial homologs, and (iv) Universal homologs. Phylogenetic analyses of these filtered gene clusters were performed in order to ensure monophyletic inheritance to exclude the possibility of horizontal inheritance of genes.

We are currently analyzing these monophyletic filtered gene cluster lists in order to form the basis of a metabolic reconstruction of the physiology and evolution of LECA. We plan to construct biochemical networks for each of the eukaryotes present in this dataset and then determine what elements of these networks have been conserved and what has arisen post-eukaryogenesis.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-337

Dynamics of seasonal adaptation in Drosophila melanogaster

Emily Behrman [1,*], Alan Bergland [2], Dmitri Petrov [3], Paul Schmidt [1]

[1]Biology, University of Pennsylvania, Philadelphia, [2]Biology, University of Virginia, Charlottesville, [3]Biology, Stanford University, Palo Alto, United States

Abstract: The rate and tempo at which populations respond to environmental change is fundamental in understanding the adaptive process. Annual seasonal rhythms produce rapid, predictable environmental changes that may result in rapid adaptation. We show that *Drosophila melanogaster* adapt rapidly and predictably to seasonal environmental changes across five years and multiple locations. Suites of complex fitness traits change in a predictable way over the 10-15 generations from spring to fall. Parallel changes in G-matrixes indicate that selection acts rapidly to alter the genetic architecture of a population. Whole genome resequencing identifies hundreds of alleles that cycle in frequency as a function of seasonal time. Functional analysis in candidate genes shows that epistatic interactions among seasonally oscillating alleles facilitate rapid adaptation by producing emergent fitness phenotypes. Together, our findings demonstrate rapid, repeatable adaptation to abiotic and biotic environmental parameters that cycle as a function of seasonal time. We show that epistatic interactions within and among genes facilitate the rapid evolutionary change that is occurring over timescales previously considered static.

Expanded summary*: My research takes a systems-level approach to understand a fundamental question in evolutionary question: the rate that adaption occurs in natural populations. While natural selection is traditionally considered to be a slow process of gradual change, I investigate the evolutionary response in natural populations of a multivoltine (multiple generations within a year) organism to the cyclic selection pressures imposed across seasonal time. This provides a novel perspective by focusing on temporal adaptation and evaluating phenotypic and genomic response in wild populations in "real time" across a 5-year long-term data set. I place my research in the context of previously described adaptation along latitudinal clines and demonstrate that the magnitude of change across seasons is equivalent to variation observed over 20º latitude. This demonstrates the strength and importance of selection across seasonal time and suggests that evolution is occurring on timescales that were previously considered static.

I take a broad perspective in understanding the dynamics of adaptation by examining the effects of both abiotic (e.g., temperature, photoperiod) and biotic (e.g., microbiota) environmental parameters. The dynamic nature and complexity of organisms motivates my whole-organism investigation of a diversity of phenotypes across seasonal time: both fitness-related traits (e.g., stress resistance and fecundity) as well as novel traits that are not traditionally considered in an adaptive framework (e.g., immune response and learning). To understand how the organism is reacting as a whole, I examine the trade-offs among theses traits through phenotypic and genetic correlations and at the molecular level through pleiotropic effects of allelic variation. Changes in both the phenotypic and genetic correlations from spring to fall counters the basic assumption of stable covariance over time and suggests that selection acts rapidly to change the genetic architecture of a population.

I take a broad perspective in understanding the dynamics of adaptation by examining the effects of both abiotic (e.g., temperature, photoperiod) and biotic (e.g., microbiota) environmental parameters. The dynamic nature and complexity of organisms motivates my whole-organism investigation of a diversity of phenotypes across seasonal time: both fitness-related traits (e.g., stress resistance and fecundity) as well as novel traits that are not traditionally considered in an adaptive framework (e.g., immune response and learning). To understand how the organism is reacting as a whole, I examine the trade-offs among theses traits through phenotypic and genetic correlations and at the molecular level through pleiotropic effects of allelic variation. Changes in both the phenotypic and genetic correlations from spring to fall counters the basic assumption of stable covariance over time and suggests that selection acts rapidly to change the genetic architecture of a population.

My research distinctively utilizes a model organism in a natural context to provide a comprehensive analysis of the relationship between phenotype, genotype and the environment that are driven by natural selection and reflect the adaptive process. I take advantage of the wealth of tools available for *Drosophila* to investigate the traits and genes involved in adaptation in the wild. Together, my research takes a holistic, whole-systems approach to understanding rapid adaptation by understanding phenotypic and genetic processes of seasonal life history evolution in response to biotic and abiotic environmental parameters.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-333

Comparing Signals of Natural Selection Between Three Native American Populations

Austin W. Reynolds [1 2,*], Jaime Mata-Míguez [2], Aida T. Miró-Herrans [2], Margarita Rzhetskaya [3], Jennifer A. Raff [4], M. Geoffrey Hayes [3 5 6], Deborah A. Bolnick [2 7]

[1]Department of Integrative Biology, [2]Department of Anthropology, University of Texas at Austin, Austin, TX, [3]Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Evanston, IL, [4]Department of Anthropology, University of Kansas, Lawrence, KS, [5]Anthropology, Northwestern University, [6]Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Evanston, IL, [7]Population Research Center, University of Texas at Austin, Austin, TX, United States

Abstract: Several studies have highlighted the adaptation of human populations to new environments worldwide. Until recently, however, studies of natural selection in Native American populations have been absent from this literature. Since humans first entered the Americas some 20,000 years ago, they have settled in many new environments across the continent. This diversity of environments has no doubt placed variable selective pressures on the populations living in each region. However the effects of these varied selective pressures have not been extensively studied to date.

In this study, we collected genome-wide data from three Native American populations in different geographic regions (Northern Alaska, the Southeastern United States, and Central Mexico). We calculated population branch statistics (PBS) and the integrated haplotype scores (iHS) to find the signals of natural selection in each population. We then compared the signals of selection across populations to explore the differences in selective pressures between populations. The differences seen between populations in our study will lay the foundation for further work on possible adaptations to the varied environments during the history and prehistory of these populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-357

Leveraging rich, multi-haplotype RAD-seq markers for population genomics in model and non-model organisms

Nicolas Rochette [1,*], Julian Catchen [1]

[1]Dpt. of Animal Biology, U. Illinois at Urbana-Champaign, Urbana, United States

Abstract: Restriction site-associated DNA sequencing (RAD-seq) is widely used in population genetics studies of non-model organisms. Importantly, when a reference genome is not available, RAD-seq can still provide genome-scale genotype data through the de novo assembly of RAD loci. However, in the absence of linkage information, the use of simple, biallelic SNP markers considerably reduces the power of biological analyses

We present a series of additions to the Stacks program suite that allow for the derivation of haplotypic markers from paired-end RAD-seq data. These results are achieved by adapting and extending several state of the art methods from different fields. In particular, the assembly of paired-end reads relies on a flexible and resilient de Bruijn graph-based algorithm, and SNP discovery uses a statistical framework proven to be robust to variable coverage. Finally, the phasing of SNPs into haplotypes uses a combination of experimental phasing (co-occurrence of SNP alleles in reads) and statistical phasing and imputation.

We demonstrate the error rates and the improvement in resolutive power granted by haplotypes over SNPs by applying the method to a simulated dataset and reanalyzing a threespine stickleback (Gasterosteus aculeatus) population from the Cook Inlet of Alaska, and also highlight some specific analyses permitted by haplotypic data.

The method is convenient to use and can generate thousands of haplotypes of potentially up to 1kb having tens of alleles, thus combining the volume of genetic markers available via SNP-related methods with the richness of microsatellite markers.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-364

Quantifying selection on codon usage in the signal peptide regions of E. coli.

Alexander Cope [1,*], Robert Hettich [1 2], Michael Gilchrist [1 3]

[1]Graduate School of Genome Science and Technology, University of Tennessee - Oak Ridge National Laboratory, Knoxville, TN 37996, [2]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, [3]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, United States

Abstract: A common feature of secreted proteins in microbes is the signal peptide. Located at the N-terminal region of a peptide chain, the signal peptide is usually characterized by a 20 to 40 amino acid sequence with a positively charged N-terminus, a hydrophobic core, and a polar cleavage site at the C-terminus. These biochemical features allow chaperones, such as SecB and SRP, to bind to the signal peptide and move the protein into position to be secreted. Previous bioinformatic analysis of signal peptides of *Escherichia coli* found this region to have the highest frequency of slow-translating codons relative to the rest of the genome. This analysis implies codons in signal peptides of *E. coli* are under a different selective pressure than the rest of the genome. To quantify differences in selection on signal peptides relative to the rest of the genome, a mechanistic model rooted in population genetics was used. This model accounts for the effects of selection, mutation bias, and genetic drift in shaping codon usage. Analysis using this framework suggests the direction of selection for codons in signal peptide regions is consistent with the rest of the genome. However, there do appear to be differences in the magnitude of selection for codons in this region. Preliminary analysis suggests codon usage in signal peptides is consistent with the N-terminal regions of genes not containing signal peptides.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS10

Tracing the evolutionary origins of an insect sexual differentiation pathway based on sex-specific alternative splicing

Judith Wexler [1,*], Artyom Kopp [2]

[1]Evolution and Ecology, [2]University of California, Davis, Davis, United States

Abstract: A diversity of genetic mechanisms control sexual differentiation in animals, but little work describes how new sexual developmental pathways evolve on large phylogenetic scales. When and how are new genes co-opted into the genetic architecture controlling male and female development? How do interactions between sexual differentiation genes evolve over time? To answer these questions, I have investigated the form and function of two key sex determination genes -- *doublesex* and *transformer* -- in a phylogenetically relevant but understudied group of insects -- hemimetabolous insects, the earlier branching outgroup to more well studied holometabolous insects. My work pinpoints when sex specific alternative splicing of the gene *transformer* evolved to the phylogenetic interval between the order Hemiptera and the order Blattodea. Using RNAi in the German cockroach, I have also demonstrated that the role *transformer* plays in female development predates this gene's sex-specific splicing. This work also shows that *transformer* regulates *doublesex* expression in female cockroaches, as it does in the well studied models of *Diptera*, *Coleoptera*, and *Hymenoptera*, although *doublesex* does not play a detectable role in female cockroach development. All together, these results are a stepwise depiction of how a well studied insect developmental pathway evolved on a long phylogenetic time scale.

Expanded summary*: Model organisms heavily skew our understanding of biology. In insects, most of what we know comes from species in the holometabolous clade -- a derived, monophyletic group (Misof, 2014). In contrast to hemimetabolous insect species, in which juveniles bear physical resemblance to adults and go through successive molts to maturity, holometabolous insects have morphologically distinct larval, pupal, and adult forms. Entomologists have long been fascinated by the ecological and evolutionary factors differentiating holometabolous organisms from the rest of insects, such as: what ecological pressures favored the evolution of a holometabolous lifestyle? What changes in developmental biology allowed for holometaboly to evolve? Why are there so many more holometabolous than hemimetabolous species? (Truman and Riddiford, 1999; Yang, 2001).

Answering these questions has been hampered by the fact that almost all model insect species (e.g., *Drosophila melanogaster, Tribolium castaneum, Bombyx mori*) are holometabolous. Learning more about the basic biology of hemimetabolous insects aids our understanding of the evolution of holometabolous organisms.

Sex-specific development in the derived holometabolous group is controlled primarily by two genes, both of which are alternatively spliced in a sex-specific manner: *doublesex* (*dsx*) and *transformer* (*tra*) (Geuvernick and Beukeboom 2014). *dsx* is a transcription factor and modulates expression of a panoply of species-specific downstream targets (Robinett, 2010); *tra* is a largely disorganized, repetitive RS-like protein necessary for female-specific splicing of *dsx*. *tra* is transcribed in both males and females, but a premature stop codon in males prevents correct translation of the gene product. In females, the protein is part of a complex which guides splicing machinery to include female specific exons in the *dsx* transcript (Verhulst, 2010).

Holometabolous insects are the only known metazoans in which sex-specific alternative splicing of a set of genes controls sexually dimorphic development. My PhD work investigates how this system in holometabolous insects arose – a question which first and foremost requires an investigation into sexual differentiation in hemimetabolous insects. In an attempt to understand when sex-specific alternative splicing of *tra* and *dsx* evolved, I used cloning and RNA sequencing to catalog the isoforms of these two genes present in both sexes of three phylogenetically relevant insect taxa – in the louse *Pediculus humanus* (order: *Pthriraptera*), the kissing bug *Rhodnius prolixus* (order: *Hemiptera*), and the German cockroach *Blattella germanica*. To understand how these genes evolved the functional roles with which we are familiar in holometabolous insects, I conducted a set of RNAi experiments knocking down both *dsx* and *tra* in the German cockroach *B. germanica*. I have discovered that *tra* evolved a role in female differentiation before it evolved the characteristic sex specific splicing patterns described in holometabolous insects. I also have demonstrated that *tra* regulated *dsx* in female insects before *dsx* evolved a role in female development. Together, these two sets of experiments reveal how gene isoform and function have evolved together over millions of years of insect evolution.

REFERENCES

Geuvernick, E. and Beukeboom, LW. (2014) Sex Dev. 8:38-49.

Misof, et al. (2014) Science. 346:763-767.
Robinett *et. al.* (2010) 8(5): e1000365
Truman, J., and Riddiford, L (1999) Nature. 401:447 - 452.
Verhulst, E. *et. al.* (2010) Curr Opin Genet Dev. 20: 376 - 383.
Yang, A. S. (2001) Evol Dev. 3(2):59-72.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-354

The performance of in-solution capture enrichment experiment for the human Y chromosome

Diana Ivette Cruz-Davalos [1,*], Alexandra Sockell [2], María Nieves-Colón [3], David G. Poznik [4], Hannes Schroeder [5], Carlos D. Bustamante [6], Anna-Sapfo Malaspinas [1], María C. Ávila-Arcos [7]

[1]Institue of Ecology and Evolution, University Of Bern, Bern, Switzerland, [2]Stanford University, California, [3]Arizona State University, Arizona, [4]23andMe, California, United States, [5]National History Museum of Denmark, Copenhagen, Denmark, [6]Stanford University, Palo Alto, United States, [7]International Laboratory for Human Genome Research, National Autonomous University of Mexico, Juriquilla, Mexico

Abstract: Uniparental markers (i.e. mitochondrial and Y chromosome DNA) are widely used in population genetics to infer the demographic history of specific lineages. Despite harbouring the largest non-recombining region in the human genome, which allows to trace the patrilineal history of populations, the Y chromosome contains large repetitive and heterochromatic stretches making it difficult to assemble good quality sequence data. This represents a bigger challenge when attempting to analyse ancient samples, where the target DNA comes in low quantities and is highly fragmented. To overcome this challenge, we implemented in-solution capture experiments targeting the Y chromosome DNA to improve the quality of the resulting assembly while reducing sequencing costs.

We compared enrichment strategies on samples excavated from the Caribbean, belonging to individuals enslaved from Africa on the 17th century as well as Puerto Rican Tainos from the Pre-Contact era. We further investigate the parameters affecting both the quality and the quantity of the data that can help us to assign the Y chromosome haplogroup. Preliminary results show that enrichment increased the depth of coverage on the Y chromosome up to 3.7 orders of magnitude, allowing us to sequence informative SNPs and learn about the paternal genetic ancestry of the individuals of interest.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-361

Phenotypic plasticity promotes recombination modification in periodic environments

Davorka Gulisija [1,*], Joshua B. Plotkin [1]

[1]University of Pennsylvania, Philadelphia, United States

Abstract: Phenotypic plasticity is known to arise in varying habitats where it diminishes harmful environmental effects. How plasticity shapes genetic architecture of traits under varying selection is unknown. Using an analytic approximation and Monte Carlo simulations, we show that balanced polymorphism and recombination modification arise simultaneously as a consequence of epistatic plastic modification in periodic environments. Under this novel finite-population scenario of recombination modification, recombination arises between a plasticity modifier locus and its target coding-locus in the absence of typically assumed antagonistic co-evolution or constant influx of mutation. Moreover, even in the absence of epistasis or initial physical linkage between the co-modified coding loci, they cluster together such that alleles with aligned effects associate in supergenes. In turn, diversity increases due to both recombination between modifier and target loci and recombination suppression between additively acting target loci. This study uncovers the role of phenotypic plasticity in the evolution of recombination rates and maintenance of supergenes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-351

LDadmix: a method to estimate LD in the ancestral populations of admixed samples

Ryan Waples [1,*], Anders Albrechtsen [1], Ida Moltke [1]

[1]Biology, University of Copenhagen, Copenhagen, Denmark

Abstract: Linkage disequilibrium (LD) is a fundamental concept in population genetics. It plays a key role in several different contexts, ranging from genetic association studies to inference of the evolutionary and demographic history of a population. It is important to be able to accurately estimate LD, and there are many viable methods for doing so. However, LD is affected by admixture, and importantly, when applied to admixed samples, existing LD estimation methods only provide estimates of post-admixture LD. This means that without unadmixed samples from a target population, existing methods cannot be used to estimate LD in that population prior to admixture and LD-based conclusions about demography and selection prior to admixture cannot be made properly. This is a problem, as many populations worthy of study are predominantly or initially found as ancestry components in admixed samples. Motivated by this issue, we present a maximum-likelihood method for estimating LD between pairs of SNPs in the ancestral source populations of a recently admixed population. The method requires only unphased genotypes and admixture proportions for a sample of individuals, and it produces estimates of haplotype frequencies in the ancestral source populations and in turn LD. First, we demonstrate that the method produces accurate estimates of LD in both simulated and real data sets. Then, we apply it to a number of recently admixed human populations to obtain new knowledge about LD in their ancestral populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OM-OS5

High-throughput pathogen detection in ancient metagenomic data

Felix M Key [1,*], Ron Huebler [1], Christina Warinner [1], Kirsten Bos [1], Wolfgang Haak [1], Johannes Krause [1], Alexander Herbig [1]

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

Abstract: Analyses of pathogens associated with ancient human remains provide insights into the human disease burden of past populations. For a long time, such analyses were limited to characteristic skeletal lesions that can be caused by only a few pathogens. Today, the increasing availability of metagenomic data obtained from ancient human remains offers a new possibility for the detection of blood borne and oral pathogens, providing direct evidence for their presence in human remains. The genomic information from ancient pathogens provides insights into their evolution and adaptation to the human host. However, existing manual approaches for pathogen detection in ancient metagenomic data are laborious and error prone. Here, we present an in-silico pipeline to automate the detection of relevant bacterial species in ancient metagenomic data. We exploit several signatures indicative of ancient pathogen presence in shallow sequenced DNA libraries. We assess the sensitivity and specificity of our pipeline and optimize parameters using simulated data sets for different source material, and with varying endogenous DNA content for over thirty relevant pathogens and commensals. Lastly, we deploy our pipeline to a collection of over 2500 ancient metagenomic libraries from human remains to infer the presence of pathogens in more than ten thousand years of human evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-365

Do bacterial species objectively exist? A case study from the genus Streptomyces.

Erik Wright [1,*], David Baum [2]

[1]Wisconsin Institute for Discovery, [2]Department of Botany, University of Wisconsin-Madison, Madison, United States

Abstract: Biologists make sense of the immense diversity of organisms by organizing them into a hierarchical taxonomy. The legitimacy of this taxonomy has been questioned for bacteria because they regularly intermix through horizontal gene transfer. Here, we analyze the genomes of 701 strains belonging to the bacterial family Streptomycetaceae, demonstrating that a hierarchical taxonomy can be established in spite of extensive horizontal gene transfer. We show that both the core-genome and pan-genome share the same majority phylogenetic history, implying that genomic makeup is dominated by vertical inheritance. Moreover, we find that about half of the clades on the consensus tree are exclusive, being composed of organisms that are more closely related to each other than to any organism outside the group. This shows that many clades persist in the face of genetic intermixing, justifying their recognition in the taxonomic hierarchy. Nonetheless, we could not detect a natural discontinuity at which to draw the species boundary, calling into doubt the existence of a natural species rank. We compare this situation to that of the genus *Saccharomyces*, a eukaryote with many genomes available and a similar genome size. Taken together, our results indicate that hierarchical structure in bacterial populations is maintained in genomes despite horizontal gene flow, suggesting that taxonomists should adopt flexible conventions for delimiting species based on societal expectations and historical precedent.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

OW-OS12

The Anopheles gambiae 1000 Genomes Project: population genomics and malaria vector control

Chris Clarkson [1,*], Alistair Miles [2], Nicholas Harding [2], Giordano Bottà [3], Dominic Kwiatkowski [2] and Ag1000g Consortium

[1]Wellcome Trust Sanger Institute, Cambridge, [2]Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom, [3]Universita di Roma, Sapenza, Italy

Abstract: The *Anopheles gambiae* 1000 Genomes Project (Ag1000G) is conducting whole-genome deep sequencing of wild-caught malaria vectors from populations across Africa with three core objectives: to discover genetic variation, describe population structure/history, and to connect these with malaria epidemiology and vector control. We describe the Ag1000G phase 2 data resource, which includes data on nucleotide variation in 1142 individual mosquitoes collected from 13 African countries. We discover more than 57 million SNPs, of which 24% are multiallelic, which corresponds to a variant allele at every 1.9 accessible bases of the genome on average. We present initial results from analyses of genetic diversity, population structure and demographic history within the phase 2 cohort, which includes several new mainland populations representing the incipient species *An. gambiae* and *An. coluzzii* and introduces two new island populations, providing a rich resource for investigating speciation processes, patterns of gene flow, and genome regions under strong selective pressures. We also present the latest results from haplotype-based analyses of recent selection at genes involved in insecticide resistance. These include several dramatic demonstrations of how insecticide pressure has caused resistance alleles to emerge independently within multiple populations, but also how some resistance-carrying haplotypes have managed to spread over vast geographical distances, including one example which has spread across the entire breadth of the African continent. The Ag1000G phase 2 data resource provides unique opportunities to study basic evolutionary biology and develop new approaches to malaria vector control.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-368

Detection and classification of hard and soft sweeps from unphased genotypes by multilocus genotype identity

Alexandre Harris [1 2,*], Nandita Garud [3], Michael DeGiorgio [1 4]

[1]Department of Biology, [2]Molecular, Cellular, and Integrative Biosciences, Pennsylvania State University, University Park, [3]Gladstone Institute, University of California, San Francisco, San Francisco, [4]Institute for CyberScience, Pennsylvania State University, University Park, United States

Abstract: Positive natural selection leads to a decrease in gene diversity at the selected site and at linked sites, producing a characteristic signature of extended haplotype homozygosity. These selective sweeps can be hard, wherein a single haplotype rises to high frequency, or soft, with multiple haplotypes simultaneously rising to high frequency. Measures of haplotype expected homozygosity have previously been used to detect selection in *Drosophila melanogaster* and humans. However, the methods as proposed require phased haplotype data, which is typically unavailable for non-model organisms, and have reduced power to detect soft sweeps due to their shallower genetic diversity. To address these limitations, we considered the application of the H12 and H2/H1 statistics of Garud et al. (2015) to unphased multilocus genotype data, denoting them as G12 and G2/G1. Forward-time simulations indicate that G12 has comparable power to H12 at detecting both hard and soft sweeps, and that G2/G1 can be used to classify hard and soft sweeps analogously to H2/H1, conditional on a genomic region having high G12 values. The reason for this power is that under random mating, the most frequent haplotypes will yield frequent multilocus genotypes. Simulations based on parameters modeling human data suggest that G12 and H12 may be best suited for detecting sweeps within 4,000 generations of sampling, and increase in power under recent population expansions. We additionally found signatures of positive selection within the LCT, HERC2, MYEF2, and SLC24A5 genes of the CEU population, corroborating and complementing previous studies.

Expanded summary*: Positive natural selection is the process by which variants conferring beneficial phenotypes increase in frequency, and is therefore a primary mechanism through which adaptation occurs. Selective sweeps, which can be hard or soft, leave behind a characteristic genomic signature which due to recombination decays proportionally to the time since selection and the strength of the selection. Hard sweeps involve a single haplotype rising to high frequency, whereas soft sweeps involve multiple haplotypes rising to high frequency. Thus, methods previously proposed to detect positive selection have focused on measuring extended haplotype homozygosity. Existing methods for detecting selective sweeps are limited in two ways. First, they typically have greater power to detect hard sweeps, which reduce genetic diversity to a greater extent than do soft sweeps. Second, to detect soft sweeps these methods require phased haplotype data, which is typically unavailable for non-model organisms. We address these limitations by applying the H12 and H2/H1 statistics of Garud et al. (2015) to multilocus genotype data, which we define as G12 and G2/G1, respectively.

Our statistics make use of the frequencies of the most frequent and second-most frequent multilocus genotypes to identify sites under selection. G12 is a measure of expected homozygosity that combines the frequency of the most and second-most frequent multilocus genotypes within a genomic window into a single value, and is elevated for both soft and hard selective sweeps. This is because the two most frequent genotypes comprise the majority of sampled genotypes in both cases. G1 (the standard measure of expected

homozygosity), however, is rarely significantly elevated for soft sweeps because no single genotype predominates. G2, in contrast, is smaller for hard sweeps than for soft sweeps because it measures the expected homozygosity with the most frequent genotype omitted. The omission of the most frequent genotype at the site of a hard sweep leaves a more diverse subsample than for a soft sweep, whose second-most frequent genotype may still predominate in the subsample. Accordingly, the ratio G2/G1 is larger for soft sweeps (and under neutrality) than for hard sweeps, and can distinguish the two given a significant G12.

We applied the G12 and G2/G1 statistics to simulated data based on parameters consistent with the human genetic literature, and to empirical data from the 1000 Genomes Project. We simulated hard sweep scenarios in which a single SNP is selected until reaching a threshold frequency between 0.1 and 1.0 (at intervals of 0.1), and soft sweep scenarios in which 2, 4, 8, 16, or 32 haplotypes bore the selected allele. We simulated three demographic scenarios: one of constant size, one consistent with the out-of-Africa bottleneck, and one consistent with recent human population expansion. We found that G12 reliably detected hard sweeps for high selection coefficients (down to threshold frequencies of 60% up to 4,000 generations before sampling), and that this power increased under a population expansion, and decreased under a bottleneck. The pattern was similar for soft sweep simulations, wherein selection scenarios with fewer selected haplotypes yielded the strongest G12 signals.

For data from the 1000 Genomes Project, we scanned all chromosomes across all SNPs in the CEU sample and calculated G12 and G2/G1. We detected regions of elevated G12 signal in expected locations, including the LCT gene of chromosome 2, as well as HERC2, MYEF2, and SLC24A5 of chromosome 15. Our application of G12 across the genome has corroborated and complemented previous studies, and we expect that additional new discoveries of positive selection will emerge from the use of our method on less well-studied systems.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-376

Convergent evolution of a modified TCA cycle in bacteria

Waldan Kwong [1,*], Hao Zheng [2], Nancy Moran [2]

[1]University of British Columbia, Vancouver, Canada, [2]University of Texas at Austin, Austin, United States

Abstract: The tricarboxylic acid (TCA) cycle is central to energy production and biosynthetic precursor synthesis in aerobic organisms. There exist few known variations of a complete TCA cycle, with the common notion being that the enzymes involved have already evolved towards optimal performance. Here, we present evidence that an alternative TCA cycle, in which acetate:succinate CoA-transferase (ASCT) replaces the enzymatic step typically performed by succinyl-CoA synthetase (SCS), has arisen in diverse bacterial groups, including among microbial symbionts of animals such as humans and insects. This finding has implications for our understanding of how central metabolic processes evolve, and their role in enabling microbial adaptations to local environmental conditions.

Expanded summary*: The tricarboxylic acid (TCA) cycle is a central metabolic pathway in aerobic organisms, playing key roles in energy production and biosynthesis. The enzymes that make up this cycle arose during early evolution and are highly conserved across all domains of life, with few known variations. Here, we present experimental evidence that an alternative TCA cycle, in which acetate:succinate CoA-transferase replaces the enzymatic step typically performed by succinyl-CoA synthetase, has arisen in diverse bacterial groups (across three phyla), including microbial symbionts of animals such as humans and insects. These include disease causing agents of humans, as well as a variety of commensal bacteria.

The repeated evolution of this modified TCA cycle may reflect adaptation to life in acetate-rich environments, such as that found in the guts of many animals. Intriguingly, our phylogenetic analysis suggests that transitions between using the classical TCA cycle pathway, and this alternative route, have occured throughout microbial evolution. The causes of these shifts are unclear, but may be driven by antagonistic interactions between the classical enzyme, succinyl-CoA synthetase, and the alternative, acetate:succinate CoA-transferase; here we also present experimental evidence supporting this. Our results illustrate the evolutionarily flexibility of a central metabolic process, and should encourage re-evaluation of the role of alternative enzymes in many bacterial species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-385

Novel assembly strategy provides data that supports ancient splits and gives new insights for the conservation of Solenodon paradoxus

Kirill Grigorev, Sergey Kliver, Alexey Komissarov, Pavel Dobrynin, Juan Carlos Martinez-Cruzado*, Alfred Roca, Taras Oleksyk

Abstract: Solenodons are insectivores found on the Caribbean island of Hispaniola and Cuba, with no surviving relatives found elsewhere. The genus is estimated to have diverged from the rest of the mammalian order ca. 78 Mya, and occupies a unique evolutionary position descending from the most ancient branching of the placental tree. Of additional interest is the conservation status of the Hispaniolan species (*S. paradoxus*) because it appears to be subdivided in two separate subspecies. The origins, the unique biology and adaptations of these enigmatic venomous species, coupled with their endangered status, can be greatly advanced given the availability of genome data. A whole genome assembly has never been previously attempted, partially due to the difficulty in obtaining high quality genome grade DNA samples from the field. This study corroborates previous estimates of ancient vicariant origins of the Solenodon lineage based on mitochondrial DNA. We hypothesized that the millions of years of island isolation made S. paradoxus genomes extremely homozygous, and took advantage of the low genetic diversity by trying out several assembly strategies using short paired end sequence libraries from multiple individuals. The string-graph based assembly strategy with Fermi software package showed the most promise compared to the standard de Bruijn approach with SOAPdenovo2; two Fermi-based pipelines yielded high quality contig assemblies, as evidenced by high N50, conserved gene content and favorable REAPR scores. An optimized Fermi-based pipeline beat the rest in gene annotation quality, providing the closest representation of the genome. This provides an argument that string graph model assembly methods may be the better choice for the homozygous genomes that that is often a hallmark of an endemic or endangered species. Once the consensus reference genome was assembled, the genome variants were called comparing the five individual sequences from the southern subspecies (*S. p. woodi*) and one sequence of the northern subspecies (*S. p. paradoxus*). Demographic histories of the two subspecies indicate a split for over at least the last 100 Kya, further suggesting that they need to be treated as separate conservation units. Genomic features of Solenodon genome were discovered and annotated, with a specific emphasis on the venomous genes that showed significant divergence from mammalian homologs, appearing to be closely related to the reptilian orthologs.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-378

Genetic structure of Tomato black ring virus population and effects of defective particles on virus replication

Beata Hasiów-Jaroszewska*, Daria Budzyńska [1], Paulina Korpys [1], Julia Minicka [1], Natasza Borodynko-Filas [1]

[1]Virology and Bacteriology, Institute of Plant Protection-NRI, Poznań, Poland

Abstract: *Tomato black ring virus* (TBRV) belongs to the *Nepovirus* genus (subgroup B) and infects a wide range of economically-important plants including tomato, potato and tobacco. The genome of TBRV consists of two single stranded RNAs of positive polarity. RNA1 is responsible for viral replication and polyproteins' processing and RNA2 for encapsidation and movement in plant. Moreover, TBRV infection can be accompanied by subviral particles such as satellite or defective RNAs (D-RNAs). Defective particles are subgenomic deletion mutants generated from the infectious virus genomes. D-RNAs might have a great impact on virus replication, accumulation and symptoms induced on plants (defective interfering RNAs). Here, we captured a snapshot of genetic variation of 32 TBRV isolates from different geographical regions and hosts. Sequence and phylogenetic analysis based on the gene encoding coat protein (CP) revealed a high level of genetic diversity among TBRV isolates in Poland and Europe. The similarity of CP gene varied from 91,5% to 99,7%. No significant correlation between CP gene variation, geographical origin, collection year and original host plant was observed. The recombinant isolates were also identified in the TBRV population. Moreover, in this study we analyzed the effects of defective particles generated from TBRV genome during prolonged passages in one host on parental virus replication. *Chenopodium quinoa, Solanum lycopersicum* and *Nicotiana tabacum* cv. Xanthi were infected with TBRV+D-RNA and TBRV-D-RNA. The symptoms and viral accumulation were monitored 7, 14, and 21 and 28 dpi. The accumulation level of TBRV was measured by qPCR using LightCycler® 96 (Roche) and statistical analyses were performed. Relative quantity of viral RNA in each sample was estimated by interpolating individual *Cq* values in the standard curve from three independent qPCR assays. The preliminary data revealed that the presence of D-RNAs has significant impact of virus accumulation and symptoms induced on infected plants, however the effect depends on the isolate and host.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-367

Relaxed selection during a recent human expansion

Stephan Peischl*

Abstract: Humans have colonized the planet through a series of range expansions, which deeply impacted genetic diversity in newly settled areas and potentially increased the frequency of deleterious mutations on expanding wave fronts. To test this prediction, we studied the genomic diversity of French Canadians who colonized Quebec in the 17th century. We used historical information and records from ~4000 ascending genealogies to select individuals whose ancestors lived mostly on the colonizing wave front and individuals whose ancestors remained in the core of the settlement. Comparison of exomic diversity reveals that i) both new and low frequency variants are significantly more deleterious in front than in core individuals, ii) equally deleterious mutations are at higher frequencies in front individuals, and iii) front individuals are two times more likely to be homozygous for rare very deleterious mutations present in Europeans. These differences have emerged in the past 6-9 generations and cannot be explained by differential inbreeding, but are consistent with relaxed selection mainly due to higher rates of genetic drift on the wave front. Demographic inference and modeling of the evolution of rare variants suggest lower effective size on the front, and lead to an estimation of selection coefficients that increase with conservation scores. Even though range expansions had a relatively limited impact on the overall fitness of French Canadians, they could explain the higher prevalence of recessive genetic diseases in recently settled regions of Quebec.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-365

The draft genome of Ruditapes philippinarum (the Manila clam), a promising model system for mitochondrial biology

Fabrizio Ghiselli*, Aleksey Komissarov, Liliana Milani, Joseph Dunham, Sophie Breton, Sergey Nuzhdin, Marco Passamonti

Abstract: The Class Bivalvia is a highly successful and ancient group including 20,000+ known species. They represent a good model for studying adaptation (anoxia/hypoxia, salinity, temperature, ...), and they are useful bioindicators for monitoring the concentration of pollutants in the water. They also make up an important source of food all over the world, with a production corresponding to ~20% of the global aquaculture yield. A striking feature of bivalves is the presence of an unusual mitochondrial inheritance system: the Doubly Uniparental Inheritance (DUI), so far detected in ~50 bivalve species. In DUI species, two mitochondrial genomes (mtDNAs) are present: one is transmitted through eggs (F-type), the other through sperm (M-type); the amino acid p-distance between conspecific M and F genomes ranges from 10% to over 50%. DUI provides a unique point of view for studying mitochondrial biology. In DUI systems: i) males are naturally heteroplasmic, with very divergent mtDNAs; ii) it is possible to study mitochondrial inheritance and bottleneck by following germ line mitochondria during development; iii) mitochondria are under selection for male functions.
Here we present the draft genome of the DUI species Ruditapes philippinarum (the Manila clam). DNA from a male individual was sequenced with 40x Illumina HiSeq and 30x PacBio RSII. The best de novo assembly was obtained with Canu assembler, with contig N50=76kb (CEGMA stats: 39.92% complete, 74.60% partial genes). Here we report the results of the first analyses and the technical challenges we faced, especially in de novo assembly.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POB-368

Genetic structure and molecular variability of Cucumber mosaic virus population from zucchini in Poland

Beata Hasiów-Jaroszewska [1], Mateusz Chrzanowski [1], Daria Budzyńska [1], Natalia Rymelska [1], Natasza Borodynko-Filas [1,*]

[1]Institute of Plant Protection-NRI, Poznan, Poland

Abstract: *Since 2012, Cucumber mosaic virus* (CMV) outbreaks of diverse symptomatology have been observed in zucchini crops in Poland. Besides isolates which displayed mosaics and chlorosis on infected plants, new necrotic isolates have also been identified. To characterize local virus populations, 27 CMV isolates of different origin and symptomatology were examined for subgroup identification, presence of satellite RNA (satRNA) and occurrence of recombination events. Sequence and phylogenetic analysis based on the genes encoding coat (CP) and movement proteins (MP) revealed that the Polish isolates belong to two subgroups: IA and II, with the prevalence of subgroup II. This study also detected four CMV isolates as natural recombinants from a subgroup IA pattern for MP gene and a subgroup II pattern for CP in RNA3. New, necrotic isolates from zucchini were found in both subgroups, and they were not recombinant variants. The analysis of purified viral RNAs revealed of necrotic isolates revealed lack of any additional RNA particles associated with genomic RNAs. The genetic structure of the virus population suggested that CMV is evolving rapidly, and new variants might appear in nature with different biological and genetic properties.
This study was funded by project No. 2014/13/B/NZ9/02108 of the National Science Centre in Poland.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Open Symposium

POA-326

Genome-wide evidence for adaptive evolution in two wild passerine species with different effective population sizes

Pádraic Corcoran [1], Toni Gossmann [1,*], Henry Barton [1], Jon Slate [1], Kai Zeng [1]

[1]Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom

Abstract: Understanding the contributions of natural selection and genetic drift to the molecular evolution of species is a long-standing goal in population genetics. Theory predicts a larger effective population size (N_e) should result in a greater efficacy of selection at the molecular level. To test this prediction, we analysed the genomes of 10 newly sequenced great tits from Europe and 10 published genomes of zebra finches from Australia. Estimates of nucleotide diversity at 4-fold sites indicates that the zebra finch has a ~3-fold larger N_e. The proportion of 0-fold substitutions fixed by positive selection (α) is high in both species (zebra finch: 64%; great tit: 48%) and is significantly higher in the zebra finch. We present the first estimates in birds for α in the untranslated regions (UTRs) (zebra finch: 43%; great tit: 33%), supporting a substantial role for adaptive changes in the UTRs. To account for the contribution of GC-biased gene conversion (gBGC), we estimated α using only nucleotides involving changes between A and T (weak-to-weak) or G and C (strong-to-strong) sites (WWSS). We find reduced estimates of α at 0-fold WWSS sites (zebra finch: 53%; great tit: 22%), suggesting that gBGC has contributed to 0-fold substitutions in both species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-6

Population genomics of Helicobacter pylori

Daniel Falush*, Kaisa Thorell, Samuel Sheppard, Koji Yahara, Elvire Berthenet

Abstract: I will describe three new features of Helicobacter pylori population biology that have been illuminated by sequencing of multiple genomes (1) genetic drift in the last 500 years during the colonization of the Americas. (2) superclones that have spread extensively within europe (3) linkage disequilibrium at selected genes due to epistasis.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-8

Genetic mapping and population genetics of an adaptive parasite trait: larval release time in schistosomes

Gabriel Mouahid [1], Frédéric D Chevalier [2], Salem Al Yafae [3], Mohamed A. Idris [4], Juliette Langand [1], Vinay Menon [2], Marina McDew-White [2], Tim J. C. Anderson[*], Hélène Moné [1]

[1]Université de Perpignan Via Domitia, Perpignan, France, [2]Genetics, Texas Biomedical Research Institute, San Antonio, United States, [3]Sultan Qaboos Hospital, Salalah, [4]Sultan Qaboos University, Muscat, Oman

Abstract: Parasites show exquisite adaptations to ensure transmission to new hosts, but the genetic bases of these adaptations are poorly understood. The timing of cercarial shedding from schistosome infected snails is critical for successful transmission to the vertebrate host. In Oman, *Schistosoma mansoni* parasites collected from rats shed cercariae nocturnally (~8pm) while parasites collected from humans shed cercariae during the day (11am). This project was designed to understand the genetic basis for nocturnal shedding. We conducted reciprocal genetic crosses between nocturnal and diurnally shedding Omani schistosomes, determined cercarial shedding profiles in parents, individual F1 and F2 progeny. We then sequenced exomes of parasites from each cross and used linkage mapping approaches to determine the genome regions underlying this trait. We found a strong quantitative trait locus on chr. 1 (LOD = 6.1) and a secondary peak on chr. 6. The chr. 1 peak contains a compelling candidate locus (Hes-1) that encodes a transcription repressor known to influence cell proliferation, embryogenesis and developmental timing in *Drosophila*. We are currently investigating the population genomics of Omani schistosome populations by exome sequencing field collected parasites showing nocturnal or diurnal shedding, and in future work, we will exploit the growing functional genetics toolbox for schistosomes to determine the loci underlying this trait and the mechanisms underlying timing of cercarial release. Our central goal is to understand, at the molecular level, a key parasite trait critical for transmission and host specificity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-11

Discovering the Opportunistic Pathogens That Increase Honeybee Mortality

Zack Shaffer [1,*], Nancy Moran [1], Kasie Raymann [1]

[1]Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: The recent decline in honeybee populations and their importance as pollinators for both native and agricultural plants have sparked a need for answers as to the cause of this collapse. Several possible causes have been proposed such as pesticides, climate change, and various pathogens. Although several pathogens of honeybees have been identified, the only two well-known bacterial pathogens are *Paenibacillus larvae* and *Melissococcus plutonius*, which exclusively cause disease in honeybee larvae. We hypothesized that other bacterial pathogens of the honeybee exist that have not been characterized yet. For, example, *Serratia* is a common opportunistic pathogen of animals, including insects and has been occasionally found in surveys of the honeybee microbiome where it is thought to be a signifier of an atypical microbiome composition. In this study, we isolated strains of *Serratia* from the guts of "sick" honeybees. We tested the virulence of *Serratia* in honeybees with normal gut microbiomes as well as in honeybees with perturbed (antibiotic treated) microbiomes and germ-free honeybees. From these tests, we were able to identify several *Serratia* strains that cause high mortality rates in honeybees. Additionally, these strains are capable of colonizing the gut and the haemolymph of honeybees. Here we present a comprehensive analysis of these newly identified *Serratia* pathogens through comparative genomics, pathogenicity tests, and virulence assays. Overall, our results suggest that *Serratia* is a common and unrecognized pathogen of honeybees, which can be highly virulent especially when the native gut microbiome is perturbed.

Statement: I am a fourth year biochemistry and microbiology double major at the University of Texas at Austin and am planning to graduate this June. I believe that attending the 2017 SMBE conference would be a great opportunity for me to learn about current research within the fields of molecular and evolutionary biology. Attending the conference would be very beneficial for me as I am preparing to continue my career in research. Exposure to new ideas and research techniques would greatly enhance my research experience and any individual research projects I might perform in the future. Presenting my findings and attending other presentations at this year's conference will allow me to develop my skills in public speaking and the presentation of scientific data, both of which are essential for a career in research. Because I plan to apply to graduate school next year, I believe that attending SMBE will give me the chance to network with potential mentors and explore the work being done at different universities and institutions. Additionally, attending the SMBE conference will give me a networking opportunity that is not typically available to most undergraduates. For example, it will give me a chance to interact with principle investigators with which I may have the opportunity to work with during my graduate studies. Finally, this award would allow me to attend my first international science conference, where I would be able to meet and interact with renowned scientists from all over the world. Thank you for your consideration.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-10

Improving the Schistosoma mansoni genome assembly using genetic crosses and linkage analysis

Frédéric D. Chevalier [1,*], Winka Le Clec'h [1], Gabriel Mouahid [2], Hélène Moné [2], Mohamed A. Idris [3], Salem Al Yafae [4], Juliette Langand [2], Nancy Holroyd [5], Alan Tracey [5], Matthew Berriman [5], Tim J. C. Anderson

[1]Genetics, Texas Biomedical Research Institute, San Antonio, United States, [2]Université de Perpignan Via Domitia, Perpignan, France, [3]Sultan Qaboos University, Muscat, [4]Sultan Qaboos Hospital, Salaleh, Oman, [5]Wellcome Trust Sanger Institute, Hinxton, United Kingdom

Abstract: The number of eukaryotic genomes sequenced is constantly increasing but assembly quality of these genomes is often poor, limiting their use for population genomic and QTL mapping studies. We describe use of laboratory genetic crosses and linkage information to assess assembly accuracy and to determine the location of unassigned scaffolds in the human blood fluke *Schistosoma mansoni*, which causes intestinal schistosomiasis in ~67 million people in Sub-Saharan Africa, Middle East and South America. The parasite genome (~363 Mb) is composed of 7 autosome and ZW sex chromosomes. The genome (v1) was released in 2009 and contained 19,022 scaffolds, while version 5 (released in 2011) contained 884 scaffolds, and v7 (unreleased), which incorporates PacBio long read sequencing and optical mapping, contains 379 scaffolds. We performed 5 experimental genetic crosses, and sequenced the ~15 Mb exome from a total of 458 F2 progeny allowing us to determine scaffold positions and identify assembly errors using segregation at ~20,000 Mendelian inherited SNPs. The maps drawn using v5 allowed us to relocate 288 unassigned scaffolds comprising 94 Mb (25% of the genome), within or between chromosomes. The maps drawn using the v7 showed a huge improvement with only 73 relocations, comprising 9.7 Mb or just 3% of the genome. Our analysis demonstrates that (i) the v7 *S. mansoni* genome assembly is in an excellent state and close to completion, (ii) PacBio and optical mapping approaches vastly improve assembly, and (iii) combining linkage with bioinformatics is a powerful approach for generating high quality genome assemblies.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-9

Robust exome sequencing of single schistosome miracidia – optimizing SNP calling and accuracy assessment

Frédéric D. Chevalier [1,*], Winka Le Clec'h [1], Tim J. C. Anderson

[1]Genetics, Texas Biomedical Research Institute, San Antonio, United States

Abstract: Small parasites and larval stages pose a problem for population genomic analyzes because limited amounts of DNA template are available, while the large size of many parasite genomes makes sequencing complete genomes prohibitively expensive. For example, schistosome adults live in human blood vessels and only microscopic larval miracidia (~150 μm long) are available for analysis. We evaluate the accuracy of exome sequencing of single miracidia following whole genome amplification and exome capture using a custom Agilent SureSelect array designed to capture 92% of the ~15 Mb exome. We used a "truth" set of 11,923 SNPs to optimize calling parameters using VQSLOD variant recalibration. We then used a test set of 16 F1 miracidia to evaluate call accuracy. Because the test set F1 miracidia were obtained from a cross between single genotype male and female worms, we could predict the SNP alleles in F1s and quantify the error rate. We scored 24,341 SNPs with an accuracy of 97.74%. We were also able to evaluate capture bias of alleles from genome regions showing high levels of polymorphism compared with the reference genome from which the capture array was designed. Extensive multiplexing of samples prior to genome capture reduced costs to ~$250/exome while maintaining accuracy. We conclude that scoring of genome-wide exomic SNPs from miracidia is feasible, economical and extremely accurate. This approach will allow research on schistosomes (and other parasite species or small invertebrates) to progress from population genetics using small numbers of markers, to population genomics utilizing genome-wide marker information.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG7

Parasite invasions: colonization and adaption of schistosomes to the new world during the slave trade

Frédéric D. Chevalier [1,*], Winka Le Clec'h [1], Philip LoVerde [2], Rafael Ramiro de Assis [3], Guilherme Oliveira [4], Safari Kinungi [5], Anouk Gouvras [6], Bonnie Webster [6], Joanne Webster [7], Aiden Emery [6], David Rollinson [6], Tim J. C. Anderson

[1]Genetics, Texas Biomedical Research Institute, [2]University of Texas Health Science Center, San Antonio, United States, [3]Centro de Pesquisas René Rachou, Belo Horizonte, [4]Instituto Technologico Vale, Belem, Brazil, [5]National Institute for Medical Research, Mwanza, Tanzania, United Republic of, [6]Wolfson Wellcome Biomedical Laboratories,, Natural History Museum, [7]Centre for Emerging, Endemic and Exotic Diseases (CEEED), Royal Veterinary College, London, United Kingdom

Abstract: Population genomic analyses of parasites such as malaria, trypanosomes and leishmania have provided insights into the selection pressures imposed by both human and invertebrate hosts, but comparable research on schistosomes has lagged behind because adult parasites live in the blood vessels and only microscopic miracidia larvae are available for analysis. We developed a robust, inexpensive approach for capturing and sequencing of the ~15 Mb *Schistosoma mansoni* exome that can be used for single miracidia hatched from eggs isolated from feces. Here we describe population genomic analysis of exome sequence data from 137 miracidia (one from each patient sampled) from Brazil, East Africa (Tanzania) and West Africa (Senegal and Niger). All the African samples come from the Schistosome Collection at the Natural History Museum (SCAN collection). We scored 155,057 SNPs and detail patterns of genomic variation in autosomal and mitochondrial DNA, characterize SNP variation at candidate vaccine and drug resistance loci, and examine geographic differentiation in allele frequencies to identify genome regions under strong directional and balancing selection. Comparisons of old and new world parasite populations are of particular interest, because parasites colonized the Americas during the slave trade where they encountered unfamiliar host snails (*Biompahalaria* spp). We observed several genome regions where variation has been purged from new world parasites. We suggest that these result from adaptation to new world snail species. In future work we will use the rapidly developing molecular tool box for schistosomes for functional analysis of the loci underlying parasite adaptation to a new continent.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-30

A FORMAL MODEL OF CLONAL EXPANSION IN BACTERIAL POPULATION GENETICS

Alice Ledda*, Xavier Didelot

Abstract: Recent advances in sequences techniques has dramatically increased the availability of bacterial sequences in the last few years.

Given the relative small cost of sequencing, it is now possible to sequence many very closely related bacteria and thus to study exhaustively their evolution at short time scales. This allows us to reconstruct the evolutionary history of their mobile elements, such as chromosomal cassettes and plasmids, and compare it with the evolutionary history of their hosts.

One phenomenon that is particularly evident in the short timescale of bacterial evolution is the so called "Clonal expansion".

Clonal expansion occurs when a certain lineage in a bacterial population has a selective advantage and thus expands, creating a clonal subpopulation.

Clearly identifying the clonal subpopulation is fundamental to show association between the mobile elemets and the clonal subpopulation.

The concept of clonal expansion is quite straightforward and widely used in literature, yet has not been formally conceptualised.

In this talk we will present a formal model of clonal expansion. We will show how this model can be used to simulate pattern resembling those observed in natural bacterial populations. We will also present a method to perform Bayesian inference under the clonal expansion model, thus allowing to estimate key parameters of our model such as effective populations size, time scales and expansion rates.
Finally we will apply this model to identify a clonally subpopulation responsible for a recent multi-resistant Staphilococcus aureus outbreak in England. We will show that the chromosomal cassette responsible for the success of this specific clone was readily lost when public health measures made it redundant.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG1

Evolutionary genetics of HIV drug resistance

Pleuni Pennings*

Abstract: Even though HIV/AIDS is still a big problem in the world, HIV drug resistance is no longer as big a problem as it once was. The reason is that treatments nowadays are very good. Most HIV positive people are therefore either on effective treatment, so that the virus is suppressed and resistance cannot evolve, or, people are not on treatment, and there is no drug pressure, so drug resistance will not evolve. So why do I still study HIV drug resistance? First of all, drug resistance in HIV is a great model system for evolutionary biology. Second, I hope that by studying HIV, I can ultimately contribute to solving drug resistance problems in other pathogens or cancer. I will talk about the lessons we have learned from studying selective sweeps, spatial structure, standing genetic variation and background selection in HIV. I will also discuss some ideas about how drug resistance evolution in HIV went from inevitable in the 1980s (with fast and very soft sweeps) to almost not occurring nowadays. Finally, I will discuss how drug resistance evolution in other pathogens compares to HIV, and whether the same solutions from HIV may work in other cases.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-29

Living with oxygen: Evolution of the unconventional O2-scaveging system in diplomonads

Alejandro Jiménez-González [1,*], Jan O. Andersson [1]

[1]Cell and Molecular Biology, Molecular Evolution, Uppsala University, Uppsala, Sweden

Abstract: Diplomonads are a group of unicellular, parasite or free-living, heterotrophic protist. They live in anaerobic or microaerophilic environments, although parasite species may experience increasing level of O_2 during infection. Diplomonads lack the traditional aerobic mitochondria. Instead, they contain either hydrogenosomes (*Spironucleus salmonicida*) or mitosomes (*Giardia intestinalis*) and energy is produced by fermentation. Most of the enzymes involved in this process are O_2-labile.

Previous studies showed that diplomonads have enzymes similar to bacterial homologous, indicative of a possible prokaryotic origin by lateral gene transfer (LGT). Here, we perform a bioinformatics study of the oxygen stress response genes in this group, with the goal to understand the evolutionary adaptation to increasing O_2-levels coupled to pathogenicity. A total of 24 enzymes involved directly or indirectly in the O_2-scaveging system are targeted. We use a phylogenetic approach to systematically investigate the origin of these enzymes.

Our results show the existence of a central eukaryotic core where different enzymes were added via LGT during the evolution of this linage. These enzymes of different origins interact to create a redox complex well-adapted for coping with changing O_2-levels. Synthesis of different thiols show different processes of LGT and gene losses in diplomonads. These processes might have played an important role in the adaptation of diplomonads to parasitic lifestyle.

Expanded summary*: Diplomonads are a group of flagellated protists which are free living or parasitic on mammals and fishes. Parasitic species have a high economic impact. For example, there are around 280 million cases of *Giardia intestinalis* infections reported annually in human. Fish parasitic species of the genus *Spironucleus* are a threat to fish farms specially of Atlantic salmon, Chinook salmon and Artic char. Diplomonads present anaerobic metabolism where some of the important enzymes involved contain oxygen-sensitive FeS cluster, e.g. pyruvate:ferredoxin oxidoreductase or ferredoxin. Previous studies have measured levels of 60 µM dissolved O_2 during infection in the upper small intestine. This concentration of O_2 can produce oxygen damage in the cell and/or deactivate important enzymes. The aim of this project was to understand the function and evolution of the O_2-scavenging system in more detail. We divided our goal, first, in the reconstruction of the O_2-scavenging system and understand better the proteins involved; and second, in the description the origin of these proteins and evolutionary differences between lineages.

Using a transcriptional profile under oxygen stress condition in *G. intestinalis*, in combination with literature studies, we could identify important genes related with the defense to this stress factor. In total, we identify 24 proteins that have a role directly or indirectly in oxygen stress condition. Previous studies have characterized individual enzymes involved in the O_2-scavenging system, mainly in *G. intestinalis*. Here we present putative oxygen-stress response pathways with complex interactions. We can divide these interactions in two main groups: proteins involved in detoxification of different reactive oxygen species (ROS); and proteins involved in repairing protein damage cause by ROS. 13 of the 24 proteins crate a complex system where O_2 and ROS are reduced to H_2O. At the same time, there are connections between this complex and the proteins involved in the repair of oxidized damage. Genomic data is available from diplomonads: *G. intestinalis, G. muris, Spironucleus salmonicida* and *Trepomonas sp. Trepomonas sp.* is a free living which has allow us to compare parasitic and free living species regarding oxygen detoxification. Using protein alignment tools, we could detect the present/absent of these 24 proteins and describe that some proteins are present in the whole group, but others are lineage, or even species, specific.

We developed a phylogenetic pipeline which we applied systematically to all 24 proteins to determine their evolutionary origin. We found that the O_2-scavenging system in diplomonads is built from a eukaryotic core and has been modified via lateral gene transfer. This modification has produced additions of new proteins with new functions. In other cases, this modification implies a replacement of proteins but the function remains. Our pipeline allowed us to detect two events of secondary lateral gene transfer between eukaryotes. The results were consistent with previous studies done in some of the proteins involved in O_2-scavenging system.

The project has provided a better understanding of how diplomonads have been adapted to microaerobic environment. The reduction of ROS in these species is not a unique linear process. Depending of the species, there are up to three different pathways to reduce O_2 and some of the steps are driven by several proteins with redundant function. Lateral gene transfer events in eukaryotes have played an

important role in this adaptation leading to a complex system. Some of the proteins that are most transcribed under oxygen stress condition, and have been considered the main barrier against ROS, have arrived to diplomonads, most likely from bacteria present in their microaerophilic environment.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG11

Divergence dating of the dispersal of Cryptococcus gattii to the North American Pacific Northwest using BEAST and RelTime.

Chandler Roe [1 2,*], Jolene Bowers [2], Shawn Lockhart [3], Crystal Hepp [1], David Engelthaler [2]

[1]School of Informatics and Computing Systems, Northern Arizona University, [2]Pathogen Genomics, Translational Genomics Research Institute, Flagstaff, [3]Mycotic Diseases Branch, CDC, Atlanta, United States

Abstract: Three genotypically distinct subtypes of the pathogenic fungus *Cryptococcus gattii* emerged in the North American Pacific Northwest (PNW) in the late 1990s and early 2000s and exhibited increased virulence and non-typical clinical manifestations. The spatial dissemination as well as temporal persistence of *C. gattii* in the temperate PNW is surprising as dogma suggested this pathogenic fungus was confined to tropical and subtropical regions. Currently, the causes and timing of introduction and dissemination of *C. gattii* in the PNW is not well understood. Using whole genome sequence data for 74 isolates collected between 1972 and 2016, we analyzed the three subtypes observed in the PNW outbreaks, VGIIa, VGIIb, and VGIIc, implementing both Bayesian coalescent and maximum likelihood methods. Tip dating was applied in both BEAST and RelTime in order to estimate evolutionary mutation rates as well as the time to most recent common ancestor (TMRCA). Molecular dating analyses suggest multiple introductions into the PNW for two of the three genotypes, and estimate TMRCA between 65 and 155 years ago for all three subtypes. Additionally, in comparing the mutation rates across the three subtypes, VGIIb's was 1.19E-8 substitutions per site per year while both VGIIa and VGIIc subtypes' was slightly faster at 1.58E-8 substitutions per site per year. High-resolution genomic methods for mutation rate calculations and estimation of introduction and dispersal-timing events aid in elucidating the evolutionary history and epidemiology of this pathogenic fungus as it disseminates in a novel environmental niche.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-1

Phylogenomic Reconstruction of the Oomycete Phylogeny Derived from 37 Genomes.

Charley McCarthy [1,*], David Fitzpatrick [1]

[1]Department of Biology, Maynooth University, Maynooth, Ireland

Abstract: The oomycetes are a class of microscopic, filamentous eukaryotes within the Stramenopiles-Alveolate-Rhizaria (SAR) supergroup which includes ecologically significant animal and plant pathogens, most infamously *Phytophthora infestans*. Single-gene and concatenated phylogenetic studies of both individual genera and the larger class have drawn conflicting conclusions for species phylogenies within the oomycetes. Supertree phylogenies combine large numbers of potentially disparate trees to determine evolutionary relationships that cannot be inferred from individual phylogenies alone, and have had successful application in many large-scale eukaryotic phylogenetic analyses. With a sufficient amount of genomic data now publically available, we have undertaken the first whole-genome phylogenetic analysis of the oomycetes using data from 37 species covering the four crown orders within the class and the majority of the phylogenetic clades with the sister taxa *Phytophthora* and *Pythium*. In our analysis, we have used both established and cutting-edge methods to generate supertree phylogenies from ubiquituous, orthologous and paralogous gene families. We also complemented these supertree phylogenies with phylogenomic network and supermatrix approaches. A genome-scale approach to oomycete phylogeny adequately resolves phylogeny for the entire oomycete class and for individual clades within *Phytophthora*. The class phylogeny for the oomycetes is also largely congruent under different methods of phylogenomic analysis, while some methods show greater resolution of relationships between individual *Phytophthora* clades. Our analysis will provide a useful backbone to large-scale phylogenomic studies of the oomycetes with a wider sampling of taxa in the future.

Expanded summary*: The oomycetes are important ecological eukaryotic pathogens of both plant and animal populations, and are ubiquitous in nature. Plant pathogenic species, including many *Phytophthora* and *Pythium* species, have had a significant impact on both agriculture and horticulture in recent human history. *Phytophthora infestans* is the causative agent of late potato blight, which was the main cause of the Great Famine in Ireland in the 1840s; the Famine lead to the death of over 1 million people in the island of Ireland and the displacement of another 1 million people through emigration to North America and Great Britain. This in turn lead to a rapid decline in the use of the Irish language in Ireland from the 1840s onward, and a gradual decline in the population of Ireland from a peak of 9 million in the 1840s down to nearly 3 million by the middle of the 20th century.

Closer to the present forest pathogens such as *Phytophthora ramorum* have caused significant declines in tree populations amongst oak forests along the west coast of North America and larch forests throughout Scotland and Ireland, as well as in nurseries and throughout the global horticulture industry. Many oomycetes also infect important crops, either as highly specialized pathogens (e.g. *Phytophthora sojae* solely infects the soybean plant *Glycine max*) or through widespread infection (e.g. many *Pythium* species are opportunistic pathogens and can spread root rot throughout farmland). Other oomycetes like *Saprolegnia* or *Aphanomyces* species, can devastate populations of juvenile fish and crustaceans such as salmon or crayfish, which can in turn impact fisheries.

Despite their impact on plant and animal ecology as well as on human activity, the oomycetes are somewhat under-represented relative to fungi or other eukaryotes in bioinformatics and genomics studies. Historically classified as fungi due to morphological and ecological similarities, modern-day phylogenomics places the class in the diverse Stramenopiles (or Heterokonts) phylum within the SAR supergroup along with both microscopic diatoms and macroscopic brown algae, which is distantly related to fungi.

The best-studied oomycete taxa like *Pythium* or *Phytophthora* are found within four "crown" orders: the marine Saprolegniales order, and the predominantly terrestrial orders Albuginales, Pythiales and Peronosporales. *Phytophthora* (order Peronosporales) is divided into 10 phylogenetic clades on the basis of multi-locus gene phylogenies, however different phylogenies have produced different conclusions on the relationships between these clades depending on the loci used. Similarly *Pythium* (order Pythiales) is divided into 10 phylogenetic clades on the basis of multi-locus gene phylogenies, but more recent analyses have suggested that the genus is polyphyletic or that *Pythium* should be amended into five new genera. Other oomycete taxa have been subject to recent revision, most notably the genus *Phytopythium* previously circumscribed as *Pythium* clade K.

A possible solution to the conflicts which may arise from different smaller-scale phylogenetic analyses is to take a genome-scale phylogenomic approach, using well-known supertree techniques such as matrix representation with parsimony (MRP) and gene tree parsimony (GTP). Using modern phylogenomic techniques, and complemented with other approaches such as supermatrix phylogenies, we have carried out a genome-scale phylogeny of the oomycetes. In our analysis, we used data from 37 complete oomycete genomes (the full extent of genomic data for the class) and 6 complete SAR genomes. The 37 oomycete genomes analysed contain species from each of the four "crown" orders, and the majority of phylogenetic clades within *Phytophthora* and *Pythium*. This is the first such analysis of the oomycetes, as was carried out using similar methodology to genome-scale analyses of fungi and other eukaryotes. Our analysis follows from a previous analysis, published in mSphere, in which we identified cases of bacterial-to-oomycete horizontal gene transfer in a number of *Phytophthora* and *Pythium* species using bioinformatics and phylogenetics techniques.

We have found that genome-scale phylogeny supports the four oomycete "crown" orders and supports many individual phylogenetic clades within those orders and their member genera. Resolving inter-clade relationships within *Phytophthora* is somewhat problematic under some methods due to under-representation of genomic data from some clades in the public domain, but MRP analysis displays relatively strong support from many inter-clade relationships within the genus. We hope that as genomic data for the oomycetes becomes more abundant our study will provide a useful template for future genomics and bioinformatics research of this important class of eukaryotic pathogens.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-4

Towards identifying Mycobacterium pinnipedii and viruses associated with Antarctic fur seals and Weddell seals

Adele Crane [1,*], Tanvi P. Honap [1], Michael Goebel [2], Anne C. Stone [3], Arvind Varsani [1]

[1]School of Life Sciences, Arizona State University, Tempe, [2]Antarctic Ecosystem Research Division, National Oceanic and Atmospheric Administration, La Jolla, [3]School of Human Evolution and Social Change, Arizona State University, Tempe, United States

Abstract: In the 2016/2017 field season, 41 buccal swabs and 12 blood samples were collected from 33 Antarctic fur seals (*Arctocephalus gazella*) and 8 Weddell seals (*Leptonychotes weddellii*) on Livingston Island in the South Shetland Islands. From these samples and more samples we will receive in the future, we will test for *Mycobacterium pinnipedii* using four qPCR assays and identify any evidence of host switching by analyzing the genome sequences of *M. pinnipedii* from positive samples. We also aim to identify and characterize known and novel viruses associated with pinnipeds using a high throughout sequencing approach and will determine their prevalence and diversity. For viruses that are of the same species, we will determine the degrees of genetic diversity, rates of intra-species recombination, and the frequency and distribution within coding regions of sites displaying evidence of selection. This research will expand our knowledge of *Mycobacterium pinnipedii* and viruses associated with Antarctic pinnipeds.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-5

Phylogenomic analysis reveals dynamic nature of karyotype evolution in of Plasmodium falciparum

Mario Alberto Cerón-Romero [1,*], Esther Nwaka [2], Zuliat Owoade [2], Laura A Katz [1 2]

[1]Department of Biology, University Of Massachusetts, [2]Department of Biological Sciences, Smith College, Northampton, United States

Abstract: *Plasmodium falciparum* is the most common causative agent of malaria in Africa and is responsible for most cases of the death of children. The genome of *P. falciparum* was sequenced in 2002 and has been extensively studied. However, many open questions remain, including understanding the chromosomal context of molecular evolutionary changes. Here we construct a phylogenomic map of all chromosomes of *P. falciparum* strain 3D7, which we built using a custom bioinformatic pipeline to analyze 2123 genes from up to 950 lineages. Inferring the phylogeny of these protein-coding sequences, we estimate the level of conservation across the chromosomes and distinguish regions that have been more involved in chromosomal rearrangements. We find 'young' regions (i.e. those with recently or fast evolving genes) in both subtelomeric and internal regions, and demonstrate that the patterns of molecular evolution in these regions differ substantially depending on their location: subtelomeric regions from different chromosomes share more paralogs than the young internal regions. Analyses of synteny demonstrate that subtelomeric regions of *P. falciparum* are actively shuffled among chromosomes, which is consistent with the hypothesis that these regions are prone to ectopic recombination (Scherf *et al.* 2001). We discuss our findings in the context of karyotype evolution of *P. falciparum*.

Expanded summary*: The success of *P. falciparum* as a parasite is due to numerous factors including its complex life cycle and genomic innovations. Studies suggest that the mechanisms of immune evasion depend on the structure of its chromosomes. *P. falciparum* has 14 chromosomes that harbor housekeeping genes in their internal regions and antigen genes at their ends. Although the functional differences among chromosome regions in *P. falciparum* has been intensely studied, very few is known about the origin and evolution of such differences. Our research is intended to fill this gap in the knowledge of *P. falciparum* by using a phylogenomic approach that allows us to study differences of chromosome structure in the framework of genome evolution of eukaryotes.

We use a phylogenomic pipeline that includes up to 950 lineages between prokaryotes and eukaryotes for mapping the phylogeny of 2123 protein coding-sequences of *P. falciparum* strain 3D7. Our phylogenomic map allows us to compare the patterns of evolution across different chromosome regions. Consistently with the literature, our phylogenomic map shows a typical pattern in which subtelomeric regions are young and internal regions are old. Additionally, with a huge dataset of paralog genes we show that subtelomeric regions of different chromosomes are actively recombining among each other but not with the internal regions, a trend that was previously predicted in studies with substantially smaller datasets of genes. Although internal regions are primarily old, they contain some young portions of 85 – 218 kb. We also tested if specifically those young portions recombine with subtelomeric regions by conducting an analysis of synteny. Our results revealed that there are not synteny blocks between subtelomeric regions and young portions of the internal regions, suggesting that there is not recombination between them.

Families of genes involved in immune evasion are primarily subtelomeric, which would promote the variation in those genes by ectopic recombination as described above. However, genes from the popular family *var*, which encode for the antigen PfEMP1, are also commonly found in internal regions of the chromosomes. Moreover, phylogenetic analysis and analysis of dN/dS show that there are not significant differences between subtelomeric and internal copies of *var* genes. This suggests that in contrast to other antigen genes such as *rif* and *stevor*, *var* genes may evolve under other mechanisms besides ectopic recombination. However, those mechanisms remain to be elucidated.

The high levels of recombination exhibited by the subtelomeric regions of *P. falciparum* have also been observed in other eukaryotic lineages and in some cases it generates important changes in the genome such as changes in the karyotype. The high recombination in subtelomeres is not always related to antigenic variation in other eukaryotic lineages, which implies that this characteristic may be older in the evolutionary history of *P. falciparum* than antigenic variation. Therefore it is necessary to study the mechanisms of antigenic variation and differences across chromosome regions of *P. falciparum* considering the genome evolution of eukaryotes.

Our work contributes to the fight against malaria by allowing a better understanding of the mechanisms of *P. falciparum* to escape the immune system. For instance, our work allows a better understanding of the link between characteristics of the pathogenicity of *P.*

falciparum and the genome evolution of eukaryotes. Furthermore, it allows demonstrating with a more reliable dataset mechanisms that were previously described such as ectopic recombination of subtelomeric regions. The better understanding of these mechanisms gives more opportunities for enhancing drug design to stop a disease that kills millions of people, particularly children.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-3

Network and Phylogenetic Analysis of Effector Proteins in the Oomycetes Reveals Species and Lineage Specific Expansions

Jamie McGowan [1,*], David Fitzpatrick [1]

[1]Genome Evolution Laboratory, Maynooth University, Kildare, Ireland

Abstract: Oomycetes are a class of fungi-like microscopic pathogens belonging to the 'Stramenopiles-Alveolata-Rhizaria' (SAR) superkingdom. They are notorious pathogens and have a devastating effect on a wide range of host species including plants, fish, fungi and mammals. Oomycetes secrete a large arsenal of effector proteins that can degrade host cell components, manipulate host immune responses and induce necrosis, enabling successful parasitic infection. Effectors can be divided into apoplastic effectors, acting in the extracellular space, and cytoplasmic effectors, which are translocated into the host cell, including the RxLR and Crinkler families of effectors.

We have investigated the expansion and evolution of effectors in 37 oomycete species, including *Albugo, Phytophthora*, *Pythium*, *Saprolegnia* and downy mildew species. We employed numerous bioinformatic techniques to identify putative secreted proteins which may be involved in pathogenicity. Our results show significant species specific expansions of effectors, including glycoside hydrolases and necrosis inducing proteins in *Phytophthora*, chitinases in *Pythium oligandrum*, and immunoglobulin cleaving proteins in *Aphanomyces astaci*. We utilised a combination of network based methods and traditional phylogenetic methods to map the evolution of these effectors. Our findings show that several effector families have undergone accelerated rates of evolution. Our network based analysis suggests that the RxLR effectors may have evolved via convergent evolution. In addition, we have identified a number of *Pythium* RxLR effectors, which *Pythium* were previously reported to lack. According to our knowledge, this is the first large scale study comparing effectors across oomycete species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-15

Avoiding real-time and evolutionary conformational flexibility to find broadly neutralizing antiviral targets in Zika and other flaviviruses

Janelle Nunez-Castilla [1 2,*], Jordon Rahaman [1], Christian A. Balbin [1], Jessica Siltberg-Liberles [1 2]

[1]Biological Sciences, [2]Biomolecular Sciences Institute, Florida International University, Miami, United States

Abstract: Zika virus (ZIKV), a close relative of the four Dengue virus (DENV) serotypes, is a flavivirus presently posing health risks to the Americas with no vaccine or anti-viral therapy yet available. DENV has long been a challenge to effective vaccine design due to antibody-dependent-enhancement (ADE) between its different serotypes. ADE is a mechanism where antibodies from one virus boost the virulence of future infections caused by closely related viruses. Given the relatedness between ZIKV and DENV, it is not surprising that ZIKV experiences ADE with DENV antibodies. Additionally, it is known that conformational flexibility in DENV contributes to ADE. Thus, DENV and other closely related flaviviruses must be considered for the development of novel vaccines and anti-viral drugs against ZIKV. Here, the polyproteins of 42 flaviviruses were analyzed in their evolutionary context to identify motifs constrained from changing amino acid identity and with low conformational flexibility, both in real-time and on evolutionary time-scales. Two such motifs were identified within the monophyletic clade containing 21 mosquito-borne flaviviruses. These motifs are located within the RNA-dependent RNA polymerase (RdRP) domain of the NS5 protein, a protein critical for replication of the viral genome. Examining shorter evolutionary time-scales, additional conserved motifs were identified for the DENV/ZIKV clade and for the West Nile virus (WNV) clade. Many of these clade-specific motifs were located in RdRP. Given the functional relevance and structural conservation of RdRP, these motifs are prime candidates for broadly neutralizing drug targets across different mosquito-borne flaviviruses, including ZIKV, DENV, and WNV.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-17

Poxviruses capture host genes during active retrotransposition in infected cells

Della Fixsen*, Kelsey Cone, Nels Elde

Abstract: Pathogens constantly adapt to continuously evolving host defenses. While much of virus evolution might be dominated by sampling point mutations, acquisition of genes via horizontal gene transfer (HGT) provides another source of genetic variation, especially among DNA viruses with relatively large genomes. Like point mutations, most HGT events may be neutral or deleterious, but some benefit virus fitness. Despite the prevalence this phenomenon in some virus lineages, the mechanism of HGT into viral genomes is not well understood. Several observations suggest retrotransposition of host mRNA as a mechanism of HGT to viral genomes, including the presence of a Short Interspersed Nuclear Element (SINE), usually mobilized by LINE-1 proteins, in gerbilpox genomes. The SINE, as well as several other poxvirus genes are flanked by signatures of retrotranposition, including target site duplications (TSDs), endonuclease sites, and poly(A) tails. To further understand the mechanisms of HGT, we devised an assay for selection of HGT events into virus genomes by modifying an established retrotransposition reporter. We find that human cells show a reduction in LINE-1 transposition upon infection with poxviruses. These results suggest that poxviruses may limit the frequency of HGT, which could protect viral genomes from "error catastrophe" by moderating the genetic influx of host genes. Our results suggest that poxviruses modulate the frequency of HGT to strike a balance between genetic diversity and fitness.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-16

Inferring viral consensus from deep sequencing data without de novo assembly

Bede Constantinides [1,*], David Robertson [1]

[1]Evolution and Genomic Sciences, University of Manchester, Manchester, United Kingdom

Abstract: The high mutation rates and short generation times of viruses permit a founding virus to establish a genetically diverse population within an infected host. Chronic viral infections in particular can therefore accumulate a great extent of sequence diversity, typically too great for use with conventional reference-based methods, demanding the use of *de novo* assembly. Yet even where read depth is ample, assemblies of viral populations such as HIV and HCV are often heavily fragmented, severely limiting their utility for evolutionary analyses in clinical and public health surveillance applications. While various dedicated viral genome assemblers are available, the hypervariable regions of viral genomes remain problematic in many cases. To facilitate consensus inference from deep sequenced viral populations we present kindel, a lightweight tool which reconciles indels present in a starting alignment through local reassembly until an optimal consensus sequence is found. By requiring an alignment to an arbitrarily appropriate reference as input, the complexities of de novo assembly are avoided and execution time considerably improved without introducing the challenge of optimal reference selection. We demonstrate the application of kindel to the assembly of hepatitis C sequences from 14 individuals in whom combined antiretroviral therapy (cART) has failed. Kindel can be used as a command line tool or Python package, and is available from the Python Package Index (https://pypi.python.org/pypi/kindel).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-414

BIOINFORMATIC ANALYSIS FOR VIRAL SEQUENCES DETECTION IN THE TRANSCRIPTOME OF FIVE SPECIES OF BATS FROM YUCATAN, MEXICO.

Diana Moreno [1,*], Carlos Machain [2], Georgina Hernandez [3], Jorge Ortega [1]

[1]Instituto Politecnico Nacional, [2]Centro de Investigaciones Regionales Dr Hideyo Noguchi, [3]Universidad Nacional Autónoma de México , Mexico City, Mexico

Poster: Bats are natural reservoirs of a wide variety of viruses, despite many of these bat-borne viruses can cause diseases in other mammals, it seems that health and fitness of bats are not reduced or affected by an evident viral disease. These observations had led to the hypothesis that bats might possess a unique and extremely variable immune system, resulted from a coevolutionary process between bats and viruses. In order to understand the evolutionary mechanism of the immune response in bats, we are trying to assess the correlation between the expression of two molecules from the adaptive immune response and the viruses that are being expressed in five species of bats. The first objective of this project is to identified which families of viruses are been expressed in our species.

RNA was extracted and sequenced with RNA-seq technology from liver samples of fifteen individuals classified in five distinct families of southern Mexico. Bioinformatic analysis revealed sequences of at least ten virus families, being *Flaviviridae, Poxviridae,* and *Retroviridae* the most frequent. Assembled contigs were homologous to functional viral transcripts or protein sequences, which are essential for virus replication, suggesting that these viruses were replicating in our bats at the moment of capture.

Key words: *Bats, bioinformatic analysis, natural reservoirs, RNA-seq, viruses.*

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POA-417

PLASMODIUM FALCIPARUM GENETIC DIVERSITY AND MALARIA TRANSMISSION: A META-ANALYSIS APPROACH USING MICROSATELLITE DATA

Tamar Carter [1,*], Lambodhar Damodaran [1], Daniel Janies [1]

[1]University Of North Carolina At Charlotte, Charlotte, United States

Poster: Characteristics of *Plasmodium falciparum* genetic diversity have been used to assess the intensity of malaria transmission. Microsatellite genotyping is one of the most common approaches to estimate the diversity of *P. falciparum* for epidemiological purposes. Early microsatellite based studies have shown that in regions where malaria transmission is high (e.g., African countries), *P. falciparum* exhibits several characters in its genetic diversity including: 1) high multiplicity of infection (MOI), defined as the percent of isolates of *P. falciparum* with more than one allele at one or more loci, 2) high allelic variation, measured as expected heterozygosity, (He) , and 2) low population structure, measured by pairwise F_{st}. In low transmission regions (e.g. South American countries), the trends in these characters are reversed. These genetic diversity estimates are useful for epidemiological surveillance in low transmission populations, where low number of cases make it difficult to track changes in parasite populations. While these trends in *P. falciparum* genetic diversity in different transmission settings have been described in general, few formal statistical analyses across studies and geography have been completed. Our study focuses on identifying factors that contribute to the variation in level and structure of *P. falciparum* genetic diversity using existing microsatellite data collected from the literature. Here we present summary statistics related to sample size, MOI, and H_e across a collection of microsatellite studies. Studies were identified using Google Scholar with the search terms "microsatellite", "falciparum", and "heterozygosity". From this search, 25 publications were identified that contained 138 separate study populations. Of these, 34 study populations included estimates that pooled data from greater than two-year collection periods and were thus excluded. Our final sample set included 104 study populations. The number of study populations from Africa, Southeast Asia, and South America were 58, 31, and 15 respectively. The mean sample size was 55.39 (SD = 51.9) across all study populations. The overall mean MOI was 42.0% (SD =30.2). Regionally, mean percent MOI was 54.2% (SD = 33.3) for Africa, 24.6% (SD = 12.9) for South America, and 28.7% (SD = 19.0) for Southeast Asia. The overall mean H_e was 0.66 (SD = 0.18). Regionally, H_e was 0.74 (SD = 0.12) for Africa, 0.35 (SD = 0.10) for South America, and 0.66 (SD = 0.12) for Southeast Asia. The overall mean MOI and H_e were as expected across regions, with the highest MOI and H_e observed in Africa where malaria transmission involving *Plasmodium falciparum* is higher than in other regions. Linear regression showed a significant association between MOI and H_e across study populations (β= 0.56, pv = 0.0007), but not when performed within specific regions (i.e. Africa, Southeast Asia, and South America). Our results show that geographic scale may affect the relationship between heterozygosity and MOI. Investigation is on-going on other genetic diversity characters including pairwise F_{st} and linkage disequilibrium in these studies. Also, future analysis will test for differences in reported genetic diversity estimates across defined high and low transmission categories.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG2

An asexual eukaryote demonstrating the Meselson Effect

Annette Macleod [1,*], W Weir [2], P Capewell [2], B Foth [3], C Clucas [2], A Pountain [4], P Steketee [2], N Veitch [2], M Koffi [5], T de Meeus [6], J Kabore [7], M Camara [8], A Cooper [2], A Tait [9], V Jamonneau [10] on behalf of 16, B Bucherton [11], M Berriman [2]

[1]College of Medicine, [2]University of Glasgow, Glasgow, [3]Sanger Institute, Cambridge, [4]Univesrity of Glasgow, Glasgow, United Kingdom, [5]Universite Jean Lorougnon GUEDE, Daloa, Côte d'Ivoire, [6]Institut de Recherche pour le Developpement, Burkina Faso, Burkina Faso, [7]Instistut de Rrecherche pour le Developpment, Montpellier, France, [8]Programme National de Bobo-Dioulasso, Conakry, Guinea, [9]University of Glagow, Glasgow, United Kingdom, [10]Institut de Recherche pour le Developpment, Monpellier, France, [11]Programme National de Lutte contre la Trypanosomiase Humaine Africaine, Conakry, Guinea

Abstract: Evolutionary theory predicts that the lack of recombination and chromosomal re-assortment in strictly asexual organisms results in a reduced ability to eliminate genomic mutations. The predicted consequence of this is that homologous chromosomes will irreversibly accumulate mutations and evolve independently, termed the Meselson effect. We applied a population genomics approach to examine this effect in an important human pathogen, Trypanosoma brucei gambiense. We determined that T.b. gambiense is evolving strictly asexually and is derived from a single progenitor, which emerged within the last 10,000 years. Thus we demonstrate the Meselson effect for the first time at the genome-wide level in any organism and show large regions of loss of heterozygosity, which we hypothesise to be a short-term compensatory mechanism for counteracting deleterious mutations. Our study sheds new light on the genomic and evolutionary consequences of strict asexuality, which this pathogen uses as it exploits a new biological niche, the human population.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG12

Tracking evolutionary dynamics over time to uncover hidden population structure in within-host Simian-HIV populations

Alison Feder [1,*], Pleuni Pennings [2], Zandrea Ambrose [3], Dmitri Petrov [4], Joachim Hermisson [5]

[1]Stanford University, Stanford, [2]Department of Biology, San Fransisco State University, San Francisco, [3]University of Pittsburgh, Pittsburgh, [4]Department of Biology, Stanford University, Stanford, United States, [5]Faculty of Mathematics, University of Vienna, Vienna, Austria

Abstract: Evolution of pathogens within the body (to the immune system, drugs, or to new hosts) is increasingly understood to be a heterogeneous process in space. Nevertheless, our understanding of this evolution is limited by our sampling ability, which is often at a single time point taken from the blood. To bridge this gap, we analyze 3330 viral sequences from four simian-HIV (SHIV) infected macaques sampled spatially (from five tissue types) and temporally over the course of infection (4-5 time points). The macaques were treated with ineffective drugs, and drug resistance emerged. The sampling reveals a dynamic process through which viral drug resistance spreads within the host, characterized by rapid establishment of multiple resistant haplotypes which spread across spatial compartments. However, we find that the population is not completely panmictic, and there is evidence of an upper limit on gene flow between compartments. We use this data to estimate the selective advantage of drug resistance (s), the population mutation rate (theta) and the migration rate between compartments. In particular, we find that although spatial compartments have different compositions, there's no evidence for local adaptation. We then generalize these techniques to show via simulations how strong selection (Ns >> 1) can reveal population structure when Nm >> 1, a regime that cannot be resolved with neutral markers. We predict that this technique will have broad applications as our ability to sample natural populations continues to improve, including implications for identifying local adaptation.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-12

Deciphering pathogenicity and antibiotic resistance islands in methicillin resistant Staphylococcus aureus genomes through gene clustering

Mehul Jani [1,*], Soham Sengupta [1], Kelsey Hu [1], Rajeev Azad [1]

[1]University of North Texas, Denton, United States

Abstract: *Staphylococcus aureus* is a versatile pathogen that is capable of infections in both humans and animals. It can cause furuncles, septicemia, pneumonia and endocarditis. Adaptation of *S. aureus* to modern hospital environment has been facilitated in part by the horizontal acquisition of drug resistance genes, such as *mecA* gene that imparts resistance to methicillin. Horizontal acquisitions of islands of genes harboring virulence and antibiotic resistance genes have made *S. aureus* resistant to commonly used antibiotics. To decipher genomic islands (GIs) in 22 hospital and 9 community associated, methicillin resistant *S. aureus* strains and classify a subset of GIs carrying virulence and resistance genes as pathogenicity and resistance islands respectively, we applied a host of methods for localizing genomic islands in prokaryotic genomes. Surprisingly, *none* of the frequently used GI prediction methods could perform well in delineating the resistance islands in the *S. aureus* genomes. Rather, a gene clustering procedure exploiting biases in codon usage for identifying horizontally transferred genes outperformed the current methods for GI detection, in particular, in delineating the boundaries of known islands in *S. aureus*, including the SCC*mec* island that harbors *mecA* resistance gene. The gene clustering approach also identified novel, yet unreported islands, with many of these found to harbor virulence and/or resistance genes. These yet unexplored islands may provide valuable information on the evolution of drug resistance in *S. aureus.*

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG5

Origin of the Mycobacterium tuberculosis Beijing family in Africa

Liliana Kokusanilwa Rutaihwa [1 2 3,*], Daniela Brites [2 3], David Stucki [2 3], Klaus Reither [2 3], Lukas Fenner [2 3], Sebastien Gagneux [2 3]

[1]Ifakara Health Institute, Bagamoyo, Tanzania, United Republic of, [2]University of Basel, [3]Swiss Tropical and Public Health Institute, Basel, Switzerland

Abstract: Human-adapted *Mycobacterium tuberculosis* complex (MTBC) comprises of 7 main phylogenetic lineages that show distinct phylogeographical distribution. Whilst some lineages are geographically restricted, others are widely spread. The latter includes Beijing family of Lineage 2.
Strains of Beijing family have recently received attention due their hypervirulence as shown in experimental and clinical settings, plus their association to drug resistance. Moreover, Beijing strains are emerging in many parts of the world including Africa. Recently, the Southeast Asian origin of the Beijing family was established. However, the origin of this family in Africa is currently unexplored. Initially, Beijing strains were reported in South Africa. From this observation we first hypothesized a single introduction of Beijing strains to South Africa directly from East Asia followed by subsequent spread to other African regions. Alternatively, there were multiple introductions associated with migration of people from Asia into different parts of the continent.
To unravel the origin(s) of Beijing strains in Africa, we used comprehensive whole genome sequence data sets of 800 Lineage 2 clinical isolates representing globally diverse collections to explore the phylogenetic relationships of African Beijing strains in the global context. Our preliminary findings suggest that Beijing strains were introduced multiple times into the continent from East Asia.

To further disentangle and confirm the multiple origins of African Beijing strains, we performed phylogeographical analyses including reconstruction of ancestral geographical states to determine the most likely origins of the Beijing strain family in different African regions. And, in addition genetic diversity analyses to support our preliminary findings.
Since human-adapted MTBC is an obligate pathogen, its phylogeographical distribution is strongly associated with human demography. Possibly, the presence of Beijing strains in South Africa is a consequence of slavery during the Dutch East Indies colony and the immigration of Chinese work-forces to the gold mines. Introductions to other African regions following movements of people from East Asia and dispersal from South Africa further facilitated the spread of Beijing strains in the continent.
Our findings give insights into the introduction and spread of Beijing family in Africa and contribute to a better understanding of the impact of the former on the dynamics of TB epidemiology.

Expanded summary*: The study will for the first time unravel the origin of Beijing family of Mycobacterium tuberculosis. The family is known for its hypervirulence as determined by different experimental and epidemiological parameters. In addition Beijing strains are associated with drug resistance as well as treatment failures. The global emergence and spread of these strains is therefore alarming. Understanding the introduction of Beijing strains in the African continent is therefore important to better grasp spread of the strain family and how it could impact the control of tuberculosis disease in the continent

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-7

Hybridization capture as a paleovirological tool for the detection of recent filoviral integrations in potential Ebolavirus hosts' genomes

Ariane Düx [1,*], Jan Gogarten [1], Linda Müller [1], Fabian Leendertz [1], Sébastien Calvignac-Spencer [1]

[1]Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Berlin, Germany

Abstract: The discovery of filovirus-like endogenous viral elements (EVEs) in mammalian genomes has unlocked an unexpected new source of information in the search for filovirus reservoirs. To date, paleovirological investigations have been limited to those species for which genomes have been partially or fully sequenced, enabling the discovery of ancient filovirus-like EVEs that are fixed in the populations of a wide variety of mammals, including rodents, bats and marsupials. We hypothesize that in species with an ongoing long-term association with filoviruses more recent EVEs may be present in parts of the population. The identification of such EVEs could help narrow down potential reservoir species. To detect recent filoviral EVEs, we tested a novel bench approach to paleovirology that uses in-solution hybridization capture with a set of RNA baits (MYbaits) spanning the genomes of the five Ebolavirus species and the two viruses within the Marburgvirus species described to date. To validate the method we spiked human DNA with plasmids containing a 292 bp fragment of the Ebolavirus L protein; we were able to detect an spiked DNA equivalent to less than 5% of individuals harboring a single heterozygous integration. Using a an initial ratio of human to filoviral DNA of approximately 10^8:1, after capture more than 16% of sequenced reads covered the Ebolavirus fragment. We then used this capture approach on libraries generated from pools of 30-90 individuals of potentially susceptible species that have been proposed as likely reservoirs: fruit bats, including *Eidolon helvum* and the three suspect reservoir species *Epomops franqueti*, *Myonycteris torquata* and *Hypsignathus monstrosus*, an insectivorous bat (*Chaerephon pumilus*), and chimpanzees (*Pan troglodytes verus*). Individuals were selected to span the species' geographic distribution and maximize genetic diversity in the pools. No filoviral EVEs were detected in any of these species or in humans from Côte d'Ivoire and the Democratic Republic of the Congo, where previous outbreaks have been reported. While these results provide no additional evidence for the role these species as Filovirus reservoirs, we nonetheless hope this method will expand the toolkit of paleovirologists. This method can be used on any species and on a large number of individuals, allowing for the detection of recent viral inserts that are rare variants within a population and does not require generating full genome sequences for a particular species of interest.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-14

Sequence of eradicated European Plasmodium falciparum genome prior to the introduction of antimalarial drugs.

Pere Gelabert [1,*], Carles Lalueza-Fox [1], Iñigo Olalde [2], Marcela Sandoval-Velasco [3], Rosa Fregel [4], Adrien Rieux [5], Carles Aranda [6], Krijn Paaijmans [7], Ivo Mueller [7], M. Thomas P Gilbert [8] and Paleogenomics Lab

[1]Institute of Evolutionary Biology (IBE) CSIC-UPF, Barcelona, Spain, [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States, [3]EvoGenomics, Natural History Museum of Denmark, University of Copenhagen,, Copenhagen, Denmark, [4]Department of Genetics, Stanford University, Stanford, California, United States, [5]Peuplements végétaux et bioagresseurs en milieu tropical (PVBMT) Laboratory, Agricultural Research for Development (CIRAD), Systèmes biologiques (BIOS) Department, St. Pierre de la Réunion, France, [6]Servei de Control de Mosquits, Consell Comarcal del Baix Llobregat, Sant Feliu de Llobregat, [7]ISGlobal, Barcelona Institute for Global Health, Hospital Clínic-Universitat de Barcelona, Barcelona, Spain, [8]EvoGenomics, Natural History Museum of Denmark, University of Copenhagen,, Copenhagen, Denmark

Abstract: Phylogenetic analysis of Plasmodium parasites has indicated that their modern-day distribution is a result of a series of human-mediated dispersal. A major question is the phylogenetic affinity of the malaria causing parasites Plasmodium vivax and falciparum in Europe. Additionaly the appearance of Plasmodium falciparum multiresistant to antimalarial therapies presents a major public health challenge in malarial endemical areas. To further explore the mechanisms that have generated this resistance and identify the genetic positions that have mutated, the complete sequence of a previous-drug implantation Plasmodium falciparum is crucial. Recently, a set of slides with blood stains of malaria-affected people from the Spain, dated between 1942 and 1944, have been found. We extracted DNA from four slides. We generated the data using Illumina sequencing after using several strategies aimed at increasing the Plasmodium DNA yield: depletion of the human genomic (g)DNA content through hybridization with human gDNA baits, and capture-enrichment using gDNA derived from P. falciparum. We have both recovered the mitochondrial and the nuclear genomes. Phylogenetic analysis of the eradicated European P. vivax mtDNA indicates that the European isolate is closely related to the present-day American haplotypes and likely entered in America post-Columbian contact. Furthermore, the European P. falciparum mtDNA and nuclear genome indicates a link with current Indian strains. The study of the drug-resistance related genes have permitted to recover the ancestral genetic variants and explore the genetic sweeps that are responsabile of drug resistance. This knowledge will contribute to an effective policy for the modelling of future drug resistances

Expanded summary*: The genus *Plasmodium* contains two of the most significant human pathogenic organisms. One, *Plasmodium vivax*, is the most widely distributed human malaria parasite outside of Africa, with a range that extends well into temperate zones, whereas the other, *Plasmodium falciparum*, is the predominant malaria parasite of humans in subsaharan Africa and the one that causes >90% malarial deaths. The current controversies on the recent evolutionary history of *P. vivax* and *P. falciparum* partly relate to the lack of genetic evidence from the European parasite that was eradicated more than half century ago. In Spain, malaria had remained endemic until the early 20th century, in particular in Andalucia, Extremadura, and Ebro delta regions; it was declared oficially eradicated in 1964. Thus, the lack of genetic evidence of the former *Plasmodium* European parasites also hampers our understanding of their biology.

We have recovered the first complete mitochondrial and nuclear genome from an eradicated strain of *P falciparum and* the complete mitochondrial genome from an eradicated *P vivax*. This achievement has two main implications. In one side we have implemented a novel technique that has permitted us to recover genetic data of Plasmodium species from infected erythrocytes present in blood samples of 80 years old. Additionally we are providing highlighted genetic data that can be used to dilucidate the biological history of P. falciparum and P. vivax and representing a turning point in the investigation of the selective sweeps responsible of the present-day

multidrug-resistance. The medical research and practice is struggling the multidrug-resistance in many areas of the word. It is estimated that this situation was the cause of 116,000 deaths in the year 2014 with a medical cost of 146 million US$.
The particular complex phylogeny of P. vivax has been mainly described by the anomalous central position of the American haplotypes, placed intermediate between Melanesian and Southeast Asian clusters nevertheless we are providing evidences of an European origin of at least the majority of the american P vivax strains. The finding of the close relation between the European and Indian P falciparum is in agreement with the position of the European isolate in the worldwide haplotype network, which is clearly unrelated to the American haplotypes
We have screened all the known genes involved in the drug-resistance mechanisms. The retrieved sequence of the European P. falciparum evidences that the presented genome carries the ancestral and sensible allele in the most crucial mutations that have been related with the resistance to: Cloroquine (Pfcrt/Pfmdr1 genes), Sulfadoxine-pyrimethamine (Pfdhps gene) and Artemisinin (Pfk13 gene). This results evidence the recent selective sweeps that have generated the multidrug-resistance P. falciparum strains. In further studies the potential of this sensible sequence must be evaluated in order to be used to identify new regions and genes that play crucial roles in the selection and adaptation of Plasmodium falciparum to malaria drug treatment.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG6

Adaptation of Baculoviruses and Nudiviruses in Drosophila and other arthropods

Tom Hill [1,*], Robert Unckless [1]

[1]Molecular Biosciences, University of Kansas, Lawrence, United States

Abstract: Immune system genes and the parasites they attempt to regulate often evolve rapidly in a molecular arms race with each other. This is especially true for viruses and the genes which suppress them. Though the molecular evolution of RNA viruses and their hosts have been extensively studied, the more complex DNA viruses have been largely ignored. DNA viruses, including baculoviruses & nudiviruses, are important insect pathogens with practical applications in population control and understanding them through an evolutionary lens is necessary for both pest control and the protection of beneficial insects. Here we look for signatures of selection across a population of baculovirus as well as a large group of baculovirus genomes to identify genes key to their activity. We also compare the frequency of a nudivirus across a set of locations and range of *Drosophila* species in hopes to identify factors associated with infection frequencies. Finally we compare population structure of viruses and their host. We find several genes show signatures of positive selection across genomes and within a population, we also find strong associations between location, species and frequency of infection. We plan to use this as a starting point to identify genes in *Drosophila* key to DNA viral immune response, as well as factors in the virus that are targets of suppression.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-21

Genomic analyses of ancient Mycobacterium tuberculosis complex strains from the Americas

Tanvi P. Honap [1,*], Åshild J. Vågene [2], Alexander Herbig [2], Michael Rosenberg [1], Jane E. Buikstra [3], Kirsten Bos [2], Johannes Krause [2], Anne C. Stone [3]

[1]School of Life Sciences, Arizona State University, Tempe, United States, [2]Max Planck Institute for the Science of Human History, Jena, Germany, [3]School of Human Evolution and Social Change, Arizona State University, Tempe, United States

Abstract: Tuberculosis (TB), caused by members of the *Mycobacterium tuberculosis* complex (MTBC), is one of the oldest human diseases. Skeletal evidence suggests that TB was prevalent in the pre-Columbian Americas; however, these pre-contact TB-causing strains were completely replaced by European *M. tuberculosis* strains after contact. Previously, we recovered genomes of three ~1000-year old human MTBC strains from coastal Peru and showed that these were closely related to strains found only in pinnipeds such as seals and sea lions. It remains unknown whether these pinniped-derived MTBC strains were responsible for the majority of the pre-contact TB cases in the Americas or if other types of MTBC strains were also present. In this study, we screened ~70 skeletal samples from pre- and post-contact era archaeological sites in the Americas for the presence of MTBC DNA using quantitative PCR assays and shotgun sequencing. Twelve samples showed adequate preservation of ancient MTBC DNA and were selected for MTBC genome enrichment and sequencing. These samples belong to the pre- and post-contact eras and are from coastal as well as non-coastal archaeological sites. We recovered 70-90% of the MTBC genomes from nine individuals, with the average coverage ranging from 5-26X. Our preliminary phylogenetic analyses show evidence for the spread of the pinniped-derived MTBC strains to non-coastal areas of the Americas possibly due to human-to-human transmission. Our data also suggest that introduction of European Lineage 4 *M. tuberculosis* strains occurred rapidly post-contact.

p { margin-bottom: 0.1in; direction: ltr; line-height: 120%; text-align: left; }

Expanded summary*: The aim of this project is to elucidate the origins of tuberculosis (TB) in the Americas using ancient DNA and a phylogenomics approach. TB has afflicted mankind for several millennia and today, it is one of the major causes of death due to a single infectious agent. In order to implement effective TB eradication strategies, it is necessary to enhance our understanding of the evolutionary history of the pathogen, the presence of nonhuman reservoirs, and the potential for zoonotic transmission.

TB is caused by bacteria belonging to the *Mycobacterium tuberculosis* complex (MTBC), which comprise a number of closely related species capable of infecting a wide range of mammalian hosts. MTBC lineages are adapted to specific hosts such as *M. tuberculosis* (humans), *M. bovis* (cattle); however cross-species transmissions are frequent. Phylogenetic studies suggest that the MTBC originated in Africa. The timing of the divergence of the MTBC is debated, but recent studies suggest it is less than 6000 years ago. The original source of introduction to humans is unknown, whereas all of the known animal-adapted MTBC lineages have evolved from a human *M. tuberculosis* lineage. Human *M. tuberculosis* lineages spread from Africa to Europe and Asia along with human population movements. The Americas remained more or less isolated after the Bering Land Bridge was submerged about 11 KYA until the arrival of Europeans in 1492. But skeletal evidence for TB in the Americas dates to about ~160 CE i.e. to the pre-European contact era, which led to the main research question of this project - how did TB originate in the pre-contact Americas?

In 2014, our group sequenced three ancient MTBC genomes from human samples from coastal Peru dating to about 1000-1200 CE. These strains were found to be distinct from any known human-adapted *M. tuberculosis* lineage and were most closely related to *M. pinnipedii*, which causes TB in seals and sea lions. Wild pinnipeds in the southern hemisphere are known to carry *M. pinnipedii* and our results suggested that pinnipeds carried MTBC strains to the coast of Peru where they were transmitted to humans who hunted seals for their meat and fur. The current study is focused on determining whether the seal-derived MTBC strains adapted to humans and spread to non-coastal parts of the Americas due to human population movements, trade routes, etc. We screened ~70 skeletal samples from individuals showing lesions characteristic of TB, extracted ancient DNA, and tested the extracts for presence of MTBC DNA using quantitative PCR assays and a shotgun sequencing approach. Twelve samples, spanning the pre- and post-contact era,

were enriched for the entire *M. tuberculosis* genome using synthetic baits and an in-solution hybridization capture protocol. We recovered nearly complete MTBC genomes from nine individuals, with mean coverage ranging from 5-26 X.

Preliminary phylogenetic analyses suggest that five post-contact era MTBC strains belong to the Euro-American human *M. tuberculosis* lineage. This lineage was brought to the Americas by Europeans and completely replaced the pre-contact era MTBC strains. Importantly, three pre-contact and one contact era MTBC strains belong to the seal-derived MTBC lineage found in the coastal Peruvian samples. The four samples are from non-coastal parts of the Americas and since these populations did not have direct contact with seals (or their products), they likely acquired the seal-derived MTBC strains from other infected humans.

The results of our study are not only important from an anthropological perspective but also inform us about the ability of this important human pathogen to infect and adapt to new hosts. Previous research had shown that a human *M. tuberculosis* lineage jumped into animals and led to the evolution of the *M. pinnipedii* lineage, but our work provides evidence for *M. pinnipedii* jumping from seals into humans and the strains then re-adapting to their human hosts. Currently, we are analysing our ancient MTBC genome data for signatures of adaptation and conducting dating analyses using BEAST to determine divergence times for the seal-derived MTBC lineage.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-24

The estimation of inter-segmental reassortment rate in influenza virus H3N2

Kangchon Kim [1], Yeongseon Park [1,*], Yuseob Kim [1]

[1]Ecoscience, Ewha Womans University, Seoul, Korea, Republic Of

Abstract: Reassortment in viruses with segmented genome is an important mechanism to drive diversity and adaptation. Previous studies have detected intra-subtype reassortment events in human influenza H3N2 by investigating between-segment incongruity in genealogies. However, the rate of reassortment, a probability that a pair of segments of an individual virus come from different parents of the previous generation, was not estimated. Here we provide the first quantitative analysis of reassortment rate in influenza H3N2. We first calculated metrics which measure incongruity in tree topology or linkage disequilibrium between hemmaglutinin (HA) and other non-antigenic segments. Next, the range of reassortment rate that reproduces the observed values was explored using simulations that are constrained to match the various patterns of H3N2 molecular evolution. Assuming one generation to be 1/80 year, within the rate between 0.001 and 0.01, the observed pattern of incongruity and linkage disequilibrium could be explained. We also confirmed that this level of reassortment is required to limit the genome-wide hitchhiking effect of positive selection on HA, thus generating much lower synonymous diversity in HA than in other segments. Reassortment rate inferred here implies that hosts may be frequently infected by multiple lineages of H3N2. We also believe that studying reassortment rate as in our work can help explain the level of differences in segmental sequence diversity in viruses with segmented genomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG3

The Stone Age Plague: 1000 years of Persistence in Eurasia

Aida Andrades Valtueña [1,*], Alissa Mittnik [1 2], Ken Massy [3 4], Raili Allmäe [5], Mantas Daubaras [6], Rimantas Jankauskas [7], Mari Tõrv, Saskia Pfrengle [2], Maria A. Spyrou [1 2], Michal Feldman [1 2], Wolfgang Haak [1 8], Kirsten I. Bos [1 2], Philipp W. Stockhammer [1 3], Alexander Herbig [1 2], Johannes Krause [1 2]

[1]Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, [2]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, [3]Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig-Maximilians-University, Munich, [4]Heidelberg Academy of Sciences, Heidelberg, Germany, [5]Archaeological Research Collection, Tallinn University, Tallinn, Estonia, [6]Department of Archaeology, Lithuanian Institute of History, [7]Department of Anatomy, Histology and Anthropology, Vilnius University, Vilnius, Lithuania, [8]School of Biological Sciences, The University of Adelaide, Adelaide, Australia

Abstract: *Yersinia pestis* is the aetiological agent of plague, a zoonotic disease associated with rodents and their fleas that has affected human populations in three historical pandemics. The discovery of molecular signatures of *Y. pestis* in Late Neolithic and Early Bronze Age (LNBA) Eurasian individuals suggests the presence of a form of plague long before the first documented pandemic (Plague of Justinian, 6th century AD).

Here, we present the first four European *Y. pestis* genomes from this period. Phylogenetic analyses show that all currently known LNBA genomes form an ancestral clade with no extant relatives. The LNBA strains expressed the now silenced *ureD* gene, which kills up to 40% of flea vectors, thus decreasing the transmission efficiency via this vector. Furthermore, these strains lack virulence factors involved in blocked-flea transmission (e.g. the *ymt* gene). Here we discuss the consequences of these findings in the light of a flea-based transmission model for plague during the LNBA.

We also exploit the time transect provided by our data to show gradual genome decay in the branch of the LNBA lineage. The loss of genes encoding flagellar proteins, which trigger host immune response, could indicate that *Y. pestis* was adapting to a new pathogenic lifestyle.

In the context of recent findings from human ancient DNA research, which suggest an increase of human mobility during this period, we propose that *Y. pestis* entered Europe from Central Eurasia during an expansion of Eurasian steppe pastoralists, persisted in Europe, and returned to Central Eurasia with subsequent human movements.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-19

Comparative genomics and transcriptomics for the identification of novel Agrobacterium tumefaciens virulence genes

Mindia Haryono [1], Shu-Ting Cho [1], Wen-Sui Lo [1], Chih-Horng Kuo [1,*]

[1]Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

Abstract: *Agrobacterium tumefaciens* is a phytopathogenic soil bacterium with extensive ranges of plant hosts and geographic distributions. One unique property of this bacterium is that it could transfer a specific segment of DNA from its tumor-inducing (Ti) plasmid to the host nuclear genome, effectively using the infected plant cells for gene expression. Based on the previous molecular characterization of virulence mechanism in the model strain C58, this bacterium has been developed into a critical tool for genetic engineering in plants. Although this technology is relatively mature, one major limitation is that the infection of many plant species and crop cultivars has remained difficult, such that the *Agrobacterium*-mediated transformation has a low efficiency or is infeasible. The recent work on the characterization of wild-type strains revealed that extensive genetic and phenotypic variations exist within the *A. tumefaciens* species complex, suggesting that a better understanding of the *A. tumefaciens* diversity could be a key to overcome this limitation on host range in current plant transformation technology. In this work, we selected a wild-type strain 1D1609 that was originally isolated from a crown gall of alfalfa (*Medicago sativa* L.) for comparative analysis with C58. Our genomic investigation revealed that these two strains share <80% of their gene content. Furthermore, the plasmid pTi1D1609 belongs to the octopine-type, which is highly divergent from the nopaline-type pTiC58. To screen for possible novel virulence genes, we conducted comparative transcriptomics analysis under the control and the virulence-induced conditions. As expected, these two strains exhibit similar patterns of up-regulation of known virulence genes belonging to the *vir* gene clusters. Intriguingly, we identified several genes with unknown function that are strongly up-regulated in 1D1609 under the induced condition and have no homolog in C58. These genes will serve as the candidates for future functional characterization by molecular genetics studies, which may then be utilized in synthetic biology for generating new *Agrobacterium* strains with higher transformation efficiencies or wider host ranges, thus improving basic biological research and agricultural biotechnology applications.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG14

Modeling the co-evolution of genes in bacterial genomes to understand the emergence of antibiotic resistance-conferring mobile elements

Florent Lassalle [1,*], Xavier Didelot [1]

[1]Infectious Disease Epidemiology, Imperial College London, London, United Kingdom

Abstract: Bacteria of the family Enterobacteriaceae are common colonizers of the human gut, but can occasionally evolve into life-threatening pathogens. Worryingly, some strains can become multi-drug resistant "superbugs", which are now considered one of the major threats to public health worldwide.

Acquiring antimicrobial resistance (AMR) through horizontal gene transfer typically involves mobile genetic elements (MGEs) such as plasmids. These MGEs often carry other 'cargo' genes with varied function, from pathogenic virulence to environmental versatility and resilience. While it is evident that exposure to antibiotics positively selects AMR genes, they also spread by hitchhiking on MGE vectors, whose evolutionary success depends on the complex interplay of diverse selective pressures, reflecting the multiple ecological adaptations encoded by MGEs.

To investigate how ecological selection for linked factors can promote or repress the spread of AMR genes, we reconstructed the complete history of gene flow amongst hundreds of enterobacteria genomes. To do so, we built a hybrid method separating inter- and intra-population evolutionary processes. We used a probabilistic phylogenetic method to reconcile the well-supported features of gene trees with the core-genome (species) tree and infer horizontal transfer events. Genes within unresolved clades were subjected to a method reconstructing the mosaic origin of polymorphisms from source populations. We integrate all single-gene history into a multi-species pan-genome evolutionary scenario to describe how MGE backbones, AMR genes and other cargo genes co-circulate and occasionally aggregate into super-pathogenic constructs.

Keywords: HGT, AMR. reconciliation

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-20

What makes a successful drug resistance allele?

Shalini Nair [1,*], Ian Cheeseman [1], Francois Nosten [2], Tim Anderson [1]

[1]Genetics, Texas Biomedical Research Institute, San Antonio, United States, [2]Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Mae Sot, Thailand

Abstract: More than 125 different *kelch13* alleles, each containing different single amino acid substitutions, have arisen in SE Asian malaria parasite (*Plasmodium falciparum*) populations under artemisinin selection over the past 15 years in a dramatic example of a soft selective event. However, just one of these alleles (C580Y) is now outcompeting other alleles in multiple different countries and is spreading towards fixation. Here we examine the transcriptional and fitness consequences of C580Y, relative to another less successful kelch13 mutation (R561H), to try to determine what is special about C580Y. Specifically, we test the hypothesis that C580Y causes less transcriptional disruption than R561H, and carries lower fitness costs. We used CRISPR/Cas9 to edit a wildtype kelch13 parasite genotype to carry C580Y, R561H, or control edits with only synonymous changes. We are now measuring transcript abundance across the parasite lifecycle, and determining relative fitness in head-to-head competition experiments, to provide direct comparisons of these two artemisinin resistance mutations on the same genetic background. Our overall goal is to better understand the dynamics of an ongoing selective event by careful functional examination of the phenotypic properties of different competing resistance alleles.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG8

Genome-wide identification of lineage and locus specific variation assoicated with pneumococcal carriage duration

John Lees [1,*], Nicholas Croucher [2], Julian Parkhill [1], Paul Turner [3 4], Stephen Bentley [1]

[1]Wellcome Trust Sanger Institute, Cambridge, [2]Department of Infectious Disease Epidemiology, Imperial College London, London, [3]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, [4]Shoklo Malaria Research Unit, Mahidol University, Mae Sot, Thailand

Abstract: The duration of asymptomatic carriage of *Streptococcus pneumoniae* in the nasopharynx is an important consideration in modelling transmission dynamics and vaccine response of a potentially life-threatening pathogen. Finding genetic effects causing variability in duration may also provide insights into the evolution of mechanisms of intra- and inter-strain competition. Existing studies are based at the serotype level only, and do not probe variation within lineages or fully quantify interactions with other environmental factors. We used longitudinal swabs taken from 598 infants combined with whole genome sequence data to determine the genetic basis for variability in carriage duration. By developing a model to estimate carriage duration from this epidemiological data, we estimate that variation in the pneumococcal genome is responsible for most (63%) of the variation in carriage length. We further quantified the variation in carriage duration caused by serotype and drug-resistance (19%) and significant locus effects (6%). A pangenome-wide association study identified 16 loci which may have an effect on carriage duration, independent of serotype. Hits at a genome-wide level of significance were mapped to prophage sequences, suggesting infection by such viruses substantially affects carriage duration. These results show that both serotype and non-serotype specific effects alter carriage duration in infants, and provide insights into pneumococcal evolution.

Expanded summary*: *Streptococcus pneumoniae* is a leading cause of invasive disease in infants, especially in low-income settings. Asymptomatic carriage in the nasopharynx is a prerequisite for disease, and its duration is an important parameter in modelling pneumococcal epidemiology. To better understand carriage, we took nasopharyngeal swabs from 598 infants every month for the first two years of life, a subset of which were whole genome sequenced. We developed a hidden Markov model for the longitudinal swab data, which utilises information from the whole cohort to account for false negative swabs. By reconstructing the most likely path through carrying and non-carrying states we estimated the carriage duration of 2157 sequenced genomes. We then performed a pangenome-wide association study (GWAS) to search for variants significantly associated with an altered carriage duration. This is the first GWAS on a continuous trait in bacteria, and is statistically highly powered. The GWAS was performed using a recently developed model which uses sequence elements as generalised variants to detect variation in the entire pangenome. By calculating the correlation in duration between close phylogenetic pairs of strains, and through modelling the distribution of effects of all variants on carriage duration, we estimated that pathogen genetics explains 63% of the variance in duration. The analysis showed significant effects on duration from serotype, erythromycin resistance, child age, previous carriage and phage infection. We found putative associations with DNA repair pathways. Previous studies on duration of pneumococcal carriage have only considered the effect of serotype, and have not been able to look at the effects of other variation with the genome, nor have they checked for carriage of unencapsulated bacteria. This study overcomes these limitations and uncovers the basis of a complex phenotype in the natural population which would otherwise be difficult to assay in lab experiments. As well as having important epidemiological corollaries, the study provides insights into the evolution of the pathogen in the nasopharynx. Additionally, recent modelling work has posited that variability in carriage duration may explain the long standing problem of how bacteria without antibiotic resistance can co-exist with resistant strains without being outcompeted. This study supports these theoretical conclusions, and provides evidence for alleles which may affect carriage duration in this way.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-28

16th century Salmonella enterica genomes from an early contact era epidemic cemetery in Mexico

Åshild J. Vågene [1,*], Michael Campana [2 3], Nelly Robles García [4], Christina Warinner [1], Maria A. Spyrou [1], Aida Andrades Valtueña [1], Daniel Huson [5], Noreen Tuross [6], Alexander Herbig [1], Kirsten I. Bos [1], Johannes Krause [1]

[1]Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany, [2]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, United States, [3]Institute of Evolutionary Medicine, University of Zurich, Zurich, Switzerland, [4]INAH, National Institute of Anthropology and History, Mexico City, Mexico, [5]Center for Bioinformatics Tübingen (ZBIT), University of Tübingen, Tübingen, [6]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, Germany

Abstract: Indigenous peoples of the New World experienced high mortality rates following contact as a result of infectious diseases, many of which were introduced by Europeans. However, most of the pathogenic agents that caused these outbreaks remain unknown and are highly debated.

The MEGAN ALignment Tool (MALT) is a high-throughput bioinformatics application used to taxonomically assign DNA reads at the genome-level. Using MALT to search for traces of ancient pathogen DNA in non-enriched sequencing data we were able to identify *Salmonella enterica* in individuals buried in an early contact era epidemic cemetery at the Mixtec site of Teposcolula-Yucundaa, Oaxaca in southern Mexico. This cemetery is linked to the 1545-1550 CE epidemic, locally referred to as *cocoliztli*, the cause of which has been debated for over a century. We applied whole-genome in-solution hybridization capture to our ancient samples in order to isolate sufficient *S. enterica* DNA for genome reconstruction.

Here we present two complete high-coverage ancient genomes for *Salmonella enterica* serovar Paratyphi C, a bacterial cause of enteric fever. In addition, we comparatively analyzed our ancient genomes in the context of modern *S. enterica* strain variation in order to assess phylogenetic relationships as well as presence and absence of virulence factors.

We propose that *S.* Paratyphi C contributed to the decline in population size sustained during the 1545 *cocoliztli* epidemic at Teposcolula-Yucundaa. This study represents a first step towards a molecular understanding of disease exchange in early contact era Mexico.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG4

Investigating Variation in the Substitution Rate Across Bacteria

Beth Gibson*, Adam Eyre-Walker [1]

[1]University of Sussex, Brighton, United Kingdom

Abstract: Rates of molecular evolution are known to vary between species and across all kingdoms of life. Here we explore variation in the rate at which bacteria accumulate mutations (substitution rates) in their natural environments. We have collated estimates of substitution rates for over 30 species of bacteria, the majority of which are pathogens evolving either within a single host or during epidemic outbreaks. Substitution rates are usually estimated by comparing sequences of isolates sampled from the wild at different time points from over a few years to several decades. Comparing across species we find large variation in these rates and therefore investigate various population genetic processes, life-history and genomic traits to elucidate why different species evolve at different rates. This includes exploring whether substitution rates are associated to mutation rates per generation, generation time, population size, the efficiency of selection/drift, recombination rates, genome size and GC content. Most strikingly we find a strong negative relationship with genome size and GC content, two genomic traits which may be connected to the mutation rate per generation, but less evidence for an association to effective populations size, the efficiency of selection or the timeframe of sampling.

Expanded summary*: Knowledge about the rates at which DNA sequences change over time is crucial to understanding how organisms evolve and adapt and how molecular evolution proceeds. Evolutionary rates (or substitution rates) are known to vary extensively across species in both prokaryotes and eukaryotes and this variation will in part be associated with species characteristics and biology. Disentangling the factors that influence evolutionary rates will help answer some fundamental questions in evolutionary biology and these have been explored in many animal and plant systems (e.g. (Bromham, 2002; Smith and Donoghue, 2008; Welch, Bininda-Emonds and Bromham, 2008; Lanfear, Welch and Bromham, 2010), but not so much in bacteria. With advances in whole-genome sequencing and bioinformatics tools a plethora of new evolutionary rate estimates have become available for many bacterial species. This is enabling researchers to investigate how bacterial populations evolve in their natural environments, with particular attention towards pathogens, both during individual host infection and epidemic outbreaks. In this study we take advantage of these substitution rate estimates and explore how and why rates vary so greatly across bacteria.

We compiled estimates of the substitution rate per year for over 30 species of bacteria from more than 60 studies. Estimates of the substitution rates of bacteria are carried out by measuring the rate at which a bacterial species accumulates mutations through time using temporarily sampled data, or concurrent samples from a population with a known origin (e.g. from fossil dates or co-speciation events) (Drummond *et al.*, 2003). These rates are measured over different timescales from a few months to over 1000 years. We find substitution rates vary by nearly 4000-fold across all species investigated.

We explore several population genetic processes, life-history and genomic traits that might influence evolutionary rates. For example, the generation time (something that has been debated in bacteria -e.g. Maughan, 2007; Weller and Wu, 2015), effective population size and the efficiency of selection (Woolfit and Bromham, 2003; Lanfear, Kokko and Eyre-Walker, 2014) and recombination rate. Some genomic traits may also be associated with substitution rates because they are correlated to other factors which impact the substitution rate. This includes genome size which is negatively correlated to the mutation rate per generation(Drake, 1991; Lynch, 2010; Sung *et al.*, 2012), a relationship between substitution rate and genome size has been observed across a broad range of microbes previously (Biek *et al.*, 2015), and GC content which might also be connected to rates of recombination (Lassalle *et al.*, 2015). We also investigate whether the time frame over which these rates are measured also influences rate variation, both across and within species, which is a widely debated topic in the literature(Rocha *et al.*, 2006; Balbi and Feil, 2007; Ho *et al.*, 2011; Biek *et al.*, 2015; Duchêne *et al.*, 2016).

Our results thus far indicate that the strongest correlates of substitution rates are genome size and GC content, two genomic traits which themselves are related. We aim to disentangle these and all of the above factors through statistical analysis and further, run our analysis in a phylogenetically controlled manner in order to account for statistical non-independence in the data. In summary, this

study will help elucidate why some bacteria evolve faster than others, and will inform our understanding of the population genetics of bacteria in general, especially in relation to pathogens.

References

Balbi, K. J. and Feil, E. J. (2007) 'The rise and fall of deleterious mutation', *Research in Microbiology*, 158(10), pp. 779–786. doi: 10.1016/j.resmic.2007.09.005.

Biek, R., Pybus, O. G., Lloyd-smith, J. O. and Didelot, X. (2015) 'Measurably evolving pathogens in the genomic era', *Trends in Ecology & Evolution*. Elsevier Ltd, 30(6), pp. 306–313. doi: 10.1016/j.tree.2015.03.009.

Bromham, L. (2002) 'Molecular clocks in reptiles: life history influences rate of molecular evolution.', *Molecular biology and evolution*, 19(3), pp. 302–309. doi: 10.1093/oxfordjournals.molbev.a004083.

Drake, J. W. (1991) 'A constant rate of spontaneous mutation in DNA-based microbes', *Proceedings of the National Academy of Sciences*, 88(August), pp. 7160–7164.

Drummond, A. J., Pybus, O. G., Rambaut, A., Forsberg, R. and Rodrigo, A. G. (2003) 'Measurably evolving populations', *Trends in Ecology & Evolution*, 18(9), pp. 481–488. doi: 10.1016/S0169-5347(03)00216-7.

Duchêne, S., Holt, K. E., Weill, F., Hello, S. Le and Holmes, E. C. (2016) 'Genome-scale rates of evolutionary change in bacteria', pp. 1–24.

Ho, S. Y. W., Lanfear, R., Bromham, L., Phillips, M. J., Soubrier, J., Rodrigo, A. G. and Cooper, A. (2011) 'Time-dependent rates of molecular evolution', *Molecular Ecology*, 20(15), pp. 3087–3101. doi: 10.1111/j.1365-294X.2011.05178.x.

Lanfear, R., Kokko, H. and Eyre-Walker, A. (2014) 'Population size and the rate of evolution', *Trends in Ecology and Evolution*. Elsevier Ltd, 29(1), pp. 33–41. doi: 10.1016/j.tree.2013.09.009.

Lanfear, R., Welch, J. J. and Bromham, L. (2010) 'Watching the clock: Studying variation in rates of molecular evolution between species', *Trends in Ecology and Evolution*, 25(9), pp. 495–503. doi: 10.1016/j.tree.2010.06.007.

Lassalle, F., Perian, S., Bataillon, T., Nesme, X., Duret, L. and Daubin, V. (2015) 'GC-Content evolution in bacterial genomes: the biased gene conversion hypothesis expands', *PLoS genetics*, 11(2), p. e1004941. doi: 10.1371/journal.pgen.1004941.

Lynch, M. (2010) 'Evolution of the mutation rate', *Trends in Genetics*. Elsevier Ltd, 26(8), pp. 345–352. doi: 10.1016/j.tig.2010.05.003.

Maughan, H. (2007) 'Rates of Molecular Evolution in Bacteria are Relatively Constant Despite Spore Dormancy'. doi: 10.1111/j.1558-5646.2007.00026.x.

Rocha, E. P., Smith, J. M., Hurst, L. D., Holden, M. T., Cooper, J. E., Smith, N. H. and Feil, E. J. (2006) 'Comparisons of dN/dS are time dependent for closely related bacterial genomes', *Journal of theoretical biology*, 239. doi: 10.1016/j.jtbi.2005.08.037.

Smith, S. A. and Donoghue, M. J. (2008) 'Rates of molecular evolution are linked to life history in flowering plants', *Science*, 322(5898), pp. 86–89. doi: 1
0.1126/science.1163197.

Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G. and Lynch, M. (2012) 'Drift-barrier hypothesis and mutation-rate evolution.', *Proceedings of the National Academy of Sciences of the United States of America*, 109, pp. 18488–92. doi: 10.1073/pnas.1216223109.

Welch, J. J., Bininda-Emonds, O. R. P. and Bromham, L. (2008) 'Correlates of substitution rate variation in mammalian protein-coding sequences.', *BMC evolutionary biology*, 8(1), p. 53. doi: 10.1186/1471-2148-8-53.

Weller, C. and Wu, M. (2015) 'A generation-time effect on the rate of molecular evolution in bacteria', *Evolution*, 69(3), pp. 643–652. doi: 10.1111/evo.12597.

Woolfit, M. and Bromham, L. (2003) ‘Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes’, *Molecular Biology and Evolution*, 20(9), pp. 1545–1555. doi: 10.1093/molbev/msg167.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG13

Haplotype analysis of complex Plasmodium falciparum infections identifies loci under selection

Angela Early [1,*], Peter Gilbert [2], Marc Lievens [3], Bronwyn MacInnis [1], Christian Ockenhouse [4], Sarah Volkman [5], Dyann Wirth [5], Daniel Neafsey [1]

[1]Broad Institute of MIT and Harvard, Cambridge, [2]Fred Hutchinson Cancer Research Center, Seattle, United States, [3]GlaxoSmithKline Vaccines, Rixensart, Belgium, [4]PATH Malaria Vaccine Initiative, Washington, [5]Harvard T.H.Chan School of Public Health, Boston, United States

Abstract: The malaria parasite *Plasmodium falciparum* possesses complex immune subversion tactics and a large antigenic repertoire whose transient expression aids immune evasion. Nevertheless, *P. falciparum* antigens bear general signatures of selection, suggesting that immune responses exert a significant selective force. Here, we use deep population sampling of natural infections to identify which specific antigenic residues are candidate immune targets. We performed amplicon-based sequencing on three antigen-coding *P. falciparum* genes: *CSP*, *TRAP*, and *SERA2*. This method provided haplotype information on all genotypes within 4400 natural infections from five African sites. At the population level, all genes showed patterns of nonsynonymous polymorphism (Pn/Ps > 1) and linkage disequilibrium that are consistent with immune-mediated balancing selection. At the intra-host level, we compared the composition of observed infections to bootstrap-simulated infections and found a non-random association of parasite haplotypes within natural infections. Non-random associations at 10 amino acid positions are maintained across all study sites, suggesting the action of selection, not drift (Dai's combined $P < 0.05$). Additionally, intra-host diversity at *SERA2* negatively correlates with patient age (-0.037 pairwise differences per year), suggesting an age-dependent accumulation of haplotype-specific immune responses. Using simulations, we estimated the selection coefficients needed to maintain both genetic patterns. Combined, these analyses point to specific amino acid combinations that appear to be jointly targeted by selection. We are currently conducting functional antibody-binding assays to investigate whether B-cell activity is the driver. Overall, these uniquely fine-grained results highlight the added value of considering both population- and host-level diversity in *Plasmodium* selection studies

Expanded summary*: In the case of several pathogens such as HIV, patterns of within-host evolution and diversity have provided insight into how the human immune system targets pathogens. Such approaches, however, have not been adopted with the malaria parasite *Plasmodium*. One reason is that *P. falciparum*, the most virulent *Plasmodium* species, causes transient rather than chronic infections, limiting our ability to detect within-host evolution. A second reason is that *P. falciparum* infections are often complex, containing two or more discrete founding lineages. Unfortunately, we do not yet possess the tools to reliably phase the genomes sequenced from such complex infections. In my research, I am addressing this gap with the use of targeted amplicon sequencing. This approach allows us to investigate evolutionary dynamics at specific loci of interest and analyze selection dynamics at both the population- and intra-host-level.

In my current project, I am studying phased sequences of three genes—CSP, *TRAP*, and SERA2—from 4400 natural infections. I treat each infection as a discrete population, which experiences both a founding bottleneck and within-population selection. Using a combination of population genetic analyses and simulations, I have found evidence for selection at specific resides within all three loci.

While all three of the antigens in this study are candidate vaccine constructs, *CSP* is of particular interest as it is used in the leading malaria vaccine, RTS,S. In phase 3 field trials, the RTS,S vaccine demonstrated lower efficacy against malaria parasites whose *CSP* sequence diverged from the vaccine construct. To improve the overall efficacy of the vaccine, researchers need to understand which specific variants mediate this differential immune effect.

My population genetic analysis of *CSP* suggests that allele-specific immune recognition of CSP may be a component of natural as well as vaccine-mediated immune protection. I am now collaborating with immunologists to functionally test these computational predictions. Currently, we are conducting antibody-binding assays to determine whether B-cell mediated immunity is driving this selection. The assays are designed to (1) test whether natural antibody binding is indeed allele-specific and (2) determine which polymorphic residues mediate the effect. One major challenge in testing polymorphic epitopes is the combinatorial complexity of

these highly diverse proteins: testing all potential combinations of amino acid variants would be experimentally intractable. For example, in our analysis of 4400 infections, we recovered 332 unique CSP haplotypes that vary at 37 amino acid positions. Testing differential antibody binding across these samples would require thousands of assays that individually target the smaller motifs within these haplotypes. Using derived allele frequency and patterns of linkage disequilibrium, I have grouped haplotypes into evolutionarily distinct "families." This has reduced the complexity of the dataset, allowing us to prioritize which haplotypes to test.
Determining the extent of allele-specific antibody-binding and pinpointing both variable and invariable targets will be key to developing more effective malaria vaccines. Further, by validating this analytical approach, I plan to demonstrate that evolutionary genetic signals provide an efficient and cost-effective way of understanding how *P. falciparum* interacts with the human immune system.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG9

Uncovering replication principles from single-cell virology experiments

Ashley Teufel [1,*], Claus Wilke [1]

[1]Department of Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: Viruses experience selective pressure on the timing and order of events during infection to maximize the number of viable offspring they produce. However, eukaryotic cells display a large amount of gene expression variability. This leads to a dynamic phenotypic landscape that viruses must face in order to produce offspring. To examine the distribution of replication dynamics displayed by viruses we fit a stochastic model to growth data from high-throughput single-cell poliovirus infection experiments. The model produces estimates of many parameters associated with intracellular dynamics including translation rates, replication rates for positive and negative strands, and the number of replication cycles between the infecting and the progeny virions. We compare the distributions of the estimates of these and other parameters across experiments performed under different drug and viral mutant conditions. We find that under some experimental conditions the distribution of the number of replication cycles between infecting and progeny virions shifts to reflect that more replication cycles occurred on average. This implies that the potential for the exploration of sequence space, as well the potential for intracellular selection among related mutant virions is enhanced under some conditions.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

OW-PG10

Stronger together: recombining bacteria can harness weak epistasis to drive adaptation

Brian Arnold [1,*], Michael Gutmann [2], Yonatan Grad [3], Samuel Sheppard [4], Jukka Corander [5], Marc Lipsitch [1], William Hanage [1]

[1]Epidemiology, Harvard T.H. Chan School of Public Health, Boston, United States, [2]Informatics, University of Edinburgh, Edinburgh, United Kingdom, [3]Immunology and Infectious Disease, Harvard T.H. Chan School of Public Health, Boston, United States, [4]Biology and Biochemistry, University of Bath, Bath, United Kingdom, [5]Biostatistics, University of Oslo, Oslo, Norway

Abstract:

The impact of epistasis on the evolution of multilocus traits depends on recombination. Population genetic theory has been largely developed for eukaryotes, many of which recombine so frequently that epistasis between polymorphisms has not been considered to play a large role in adaptation and has been compared to the fleeting influence of non-heritable effects. Many bacteria also recombine, some to the degree that their populations are described as 'panmictic' or 'freely recombining'. However, whether this recombination is sufficient to limit the ability of selection to act on epistasis is unknown. We create a sensitive method to quantify homologous recombination in five bacterial pathogens and use these parameter estimates in a multilocus model of bacterial evolution with additive and epistatic effects. We find that even for highly recombining species (e.g. *Streptococcus pneumoniae* or *Helicobacter pylori*), selection may act on the cumulative effects of weak interactions between distant mutations. Furthermore, whether selection acts more efficiently on physically proximal loci depends on the average recombination tract length. Epistasis may thus play an important role in the adaptive evolution of bacteria and, unlike in eukaryotes, does not need to be strong, involve near loci, or require specific metapopulation dynamics.

Expanded summary*:

Epistasis for fitness traits may arise from any nonlinearity in the multi-locus genotype-to-fitness map. The role of epistasis in adaptive evolution has been debated since the origin of population genetics, with Ronald Fisher and Sewall Wright having opposing views. This debate has been shaped by applications to sexually reproducing eukaryotes, in which crossover recombination and independent assortment of chromosomes rapidly destroy allele combinations, thus opposing selection on particular multi-locus genotypes. As a result, the effect of individual alleles is largely the average over all genetic backgrounds, since background-specific effects are transient. Exceptions to this paradigm arise only when epistatic effects are very strong, interacting loci are tightly linked, or specific metapopulation conditions (proposed by Wright) limit the ability of recombination to dominate microevolutionary processes (but evidence for these conditions in nature is limited). Consequently, many evolutionary biologists have followed the views of Fisher and have not given epistasis a prominent role in adaptation theory.

However, the majority of life on earth does not sexually reproduce. Bacteria, which have colonized almost every conceivable ecological niche, also recombine to varying degrees through multiple mechanisms, leading to a continuum of population structures from clonal to "fully sexual". Yet in bacteria two factors create more genome-wide linkage disequilibrium (LD): (1) homologous recombination events (the dominant mode of genetic exchange) usually involve only small DNA segments, affecting few genomic loci and (2) bacteria typically have one main circular chromosome and thus have no analog of independent assortment to break down LD. The most widely studied population genetic models, those for eukaryotes, have not been directly applicable to bacteria (or Archaea), and consequently we know little about how these factors affect adaptive processes and the evolution of multi-locus phenotypes such as antibiotic resistance, antigenic profile, or metabolic output, all of which likely contain epistatic interactions. An important question is then whether highly recombining bacteria decouple interacting mutations frequently enough to prevent selection on weak interaction effects. It also remains unknown if, as in eukaryotes, selection acts more efficiently on physically close epistatic interactions that are likely co-inherited. Answering these questions, which are relevant for the entire bacterial and Archaeal kingdoms of life, requires multi-locus models with epistasis that account for these features of bacterial evolution, together with accurate estimates of genome-wide recombination parameters.

We develop a novel method that uses Approximate Bayesian Computation and machine-learning to estimate recombination parameters from genomic data of five bacterial pathogens: *Staphylococcus aureus*, *Campylobacter jejuni*, *Streptococcus pneumoniae*, *Neisseria gonorrhoeae*, and *Helicobacter pylori*. These bacteria exhibit dramatically different degrees of genome-wide LD and include some of the most highly recombining bacteria known. We then use these recombination parameter estimates in a multi-locus model of bacterial evolution to study selection on standing genetic variation when both additive and epistatic effects contribute to fitness differences between individuals. In contrast with eukaryotes, we find that in bacteria selection may act on the cumulative effects of weak epistatic interactions ($N|s| \approx$ 1-10) regardless of their physical proximity on a circular chromosome, even for highly recombining pathogens that have been previously labeled as "freely recombining". Thus, while recombination is sufficiently strong in many bacteria to destroy phylogenetic signal in gene trees and to prevent periodic selective sweeps from purging genome-wide variation, it is not capable of hindering selection on epistatic interactions between polymorphisms.

This work has broad importance for the population genetics of microbes. Interest in sexual eukaryotes has inspired the simplifying convention of gene-centered models that promote the importance of additive gene effects and ignore epistasis. This gene-centric view pervades discussions of evolution and adaptation, including those on bacteria, many of which recombine extensively and are commonly reduced to "core" and "accessory" genes, with niches providing selective advantages to specific genes. However, we show this powerful model quickly breaks down for recombining bacteria in the presence of weak epistasis, even for distantly spaced loci in highly promiscuous species. Thus, epistasis between polymorphisms may play a critical role in adaptive processes across a majority of the tree of life, and unlike eukaryotes, these interactions do not need to be strong or physically close, and do not require specific metapopulation dynamics to permit efficient selection.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-23

Analysis of Plasmid Evolution Using Networks

Ignacio Riquelme Medina [1,*], James McInerney [2], Chris Knight [3]

[1]School of Biological Sciences, [2]Division of Evolution & Genomic Sciences, [3]School of Earth and Environmental Sciences, University of Manchester, Manchester, United Kingdom

Abstract: Plasmids are mobile genetic elements that are present in organisms like prokaryotes or fungi. They can be transmitted between individuals (horizontal gene transfer) and frequently provide evolutionarily and functionally advantageous traits to their host cell, including antibiotic resistance or toxin production. Because plasmids are usually not essential components of the genome, and because they play a significant role in horizontal gene transfer, their evolutionary histories are expected to be different to the rest of the genome. In addition, there are no gene families that are found in all plasmids, making standard phylogenomic analysis impossible. In this project we have examined the evolutionary history of plasmids using sequence similarity and bipartite network approaches. We outline how plasmid networks differ from genomic networks. We characterize the plasmid networks and we identify factors that play a role in shaping these networks.

Expanded summary*: Plasmids are small DNA molecules that are present in organisms like bacteria or fungi, that can be transmitted between individuals of the same generation (horizontal gene transfer) and generally give advantageous traits to the individual like antibiotic resistance or toxin production. Because plasmids are not part of the genomic DNA and play a big role in the horizontal gene transfer their evolution is expected to have some differences compared to that of the genomic DNA. In order to better grasp how the plasmid evolution is shaped and what are its main factors we will be analysing several hundreds of plasmids sequences and using network based methods amongst other methods to better learn how the plasmids sequences interact, trying to infer how they evolved and the main factors that affect this evolution.
If this project produces significant results we will have a better understanding of how the horizontal gene transfer influences evolution and have a better idea of what factor influence plasmids evolution and how these factors differ from the genomic DNA evolution. These results could give us a better understanding of how antibiotic resistance spreads, could help us in the development of Sequence similarity networks in evolution and might have implications in synthetic biology and the possibility of making new kinds of mobile genetic elements.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-26

Spatial and temporal genomic diversity of Plasmodium falciparum in Southeast Asia provides insight into parasite migration patterns

Amol Shetty [1 2,*], Christopher Jacob [3], Sonia Agrawal [2 4], Fang Huang [4], David Saunders [5], Chanthap Lon [6], Pascal Ringwald [7], Kay Thwe Han [8], Tin Maung Hlaing [9], Myaing Nyunt [4], Joana Silva [1], Kathleen Stewart [10], Christopher Plowe [4], Shannon Takala Harrison [4], Timothy O'Connor [1 11] and Artemisinin Resistance Confirmation, Characterization, and Containment (ARC3), Artemisinin Resistance Containment and Elimination (ARCE), MalariaGEN Community Project

[1]Institute for Genome Sciences, [2]Epidemiology and Human Genetics, University of Maryland School of Medicine, Baltimore, Maryland, United States, [3]Welcome Trust Sanger Institute, Hinxton, United Kingdom, [4]Division of Malaria Research, Institute for Global Health, University of Maryland School of Medicine, Baltimore, Maryland, United States, [5]Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand, [6]Armed Forces Research Institute of Medical Sciences, Phnom Penh, Cambodia, [7]Global Malaria Programme, World Health Organization, Geneva, Switzerland, [8]Department of Malaria Research, Ministry of Health, Yangon, [9]Defence Services Medical Research Centre, Naypyitaw, Myanmar, [10]Center for Geospatial Information Science, University of Maryland College Park, College Park, Maryland, [11]Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, Maryland, United States

Abstract: Estimates of *P. falciparum* migration are integral to understanding the spread of malaria and to develop containment and elimination measures. We examined segments of the parasite genome that are identical-by-descent (IBD) to elucidate fine-structure patterns of parasite populations influenced by malaria transmission or selective sweeps of drug resistance and to infer geographical patterns of migration across Southeast Asia. The shared IBD haplotypes were determined using a *P. falciparum* DNA microarray for 2126 samples collected in ~60 districts from 27 provinces of Africa and Asia. Admixture analyses revealed 9 ancestral populations, one each for isolates in Africa, China, Bangladesh, and eastern provinces of the Greater Mekong Subregion and 5 from isolates in Myanmar, Thailand, and Cambodia. We observed increased IBD sharing between sites in close geographic proximity and also between some geographically distant sites in Cambodia, which may reflect migration through human movement. Segregation by IBD tract length illustrates varying spatial patterns of relatively recent migration in Cambodia. Patterns of IBD sharing based on relatedness between isolates sharing kelch13 gene haplotypes mirror patterns of spreading artemisinin drug resistance. IBD sharing between samples stratified as non-admixed and admixed across populations provide insights into the directionality of parasite migration. IBD estimation is a promising method to understand migratory patterns of the parasite.

Expanded summary*: As stated by the World Health Organization (WHO) report, malaria continues to pose a significant health problem despite multiple efforts to stem its spread. The WHO report estimated 214 million new cases of malaria and 438,000 deaths across 91 countries worldwide in 2015. Although there has been an overall decrease in the incidence and number of deaths since 2000, most cases and deaths have occurred in the African and Southeast Asian regions. Antimalarial drugs have helped reduce malarial incidence and deaths. However, prolonged use of these drugs coupled with genetic diversity of the parasite has resulted in the development of multi-drug resistance, limiting treatment options. Artemisinin resistant *P. falciparum* has shown increased incidence in provinces from Cambodia, Myanmar, and neighboring provinces from Laos, Thailand, and Vietnam (comprising the Greater Mekong Subregion, GMS). Single nucleotide polymorphisms in the kelch13 (*K13*) gene associated with artemisinin resistance in the GMS were shown to spread between countries as well as emerge from de novo polymorphisms within regions. Hence, knowledge of both emergence and spread of drug resistance can guide containment and elimination interventions. Here, we use identity-by-descent

(IBD) methods to detect population structure and migratory patterns of *P. falciparum*, which provide insight into its emergence and spread. We use IBD sharing to identify potential sources and sinks of migration as well as the spatial patterns in the spread of drug resistance. We observed increased regional relatedness within proximal districts and between geographically distant districts across the Myanmar-Thailand border and within Cambodia. When segregated by IBD tract lengths, we observed changes in patterns of migration over time which may reflect recent migration through human movement. We estimated the spread of artemisinin resistance within Cambodia based on *K13* haplotypes to identify multiple de-novo occurrences of artemisinin resistance which rapidly spread within and across different Cambodian provinces. Overall, examination of shared IBD haplotypes is a promising method for delineating contemporary patterns of parasite migration that can be used to identify sources and sinks of malaria transmission. Understanding these estimates and migratory patterns of parasite would help alleviate the spread of malaria within the GMS and develop potential measures to stem the spread of drug resistance in Southeast Asia.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Pathogen genomics

POB-27

Inferring determinants of transmissibility from viral sequence data

Casper Lumby [1], Chris Illingworth [1,*]

[1]Department of Genetics, University of Cambridge, Cambridge, United Kingdom

Abstract: Viruses causing acute infection exist under evolutionary pressure both to grow rapidly within a single host, and to transmit to found infections in new hosts. Transmission events can be studied by collecting short-read data from hosts before and after the event, however there are multiple challenges to identifying variants under selection for transmissibility. Firstly, the population bottleneck acting upon the virus during transmission may distort the composition of the viral population, such that genetic drift must be distinguished from selection for transmissibility. Secondly, traditional population genetic methods for inferring bottleneck size describe sequences at the haplotype level; given short-read sequence data haplotypes may only be inferred via reconstruction. Thirdly, short-read data gives a noisy representation of the actual viral populations before and after transmission. We describe a novel solution to these challenges, and demonstrate its application to identify genetic variants responsible for increased viral transmissibility using data collected from influenza transmission studies.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

POA-146

Functional and evolutionary characterization of plant gene families that duplicate preferentially through genome duplication

Setareh Tasdighian [1 2], Michiel Van Bel [1 2], Zhen Li [1 2], Yves Van De Peer [1 2], Lorenzo Carretero-Paulet [1 2], Steven Maere [1 2,*]

[1]Department of Plant Biotechnology and Bioinformatics, Ghent University, [2]Center for Plant Systems Biology, VIB, 9052 Gent, Belgium

Abstract: Gene duplication is a major mechanism providing raw material for evolution to act upon. One of the research foci in the gene duplication field is to study the relative impact of different modes of gene duplication, in particular whole-genome multiplication (WGM) versus small-scale duplication (SSD), on the gene complement of an organism. In several lineages, it has been observed that particular functional categories of genes, such as transcription factors, signal transducers and complex-forming genes, are preferentially retained after WGM while they rarely duplicate through SSD, a pattern referred to as reciprocal retention. This peculiar duplication behavior is hypothesized to stem from dosage balance effects: WGM, in contrast to SSD, is thought to generally preserve the stoichiometric balance between interacting gene products, entailing that dosage balance-sensitive genes should preferentially duplicate through WGM. However, strong evidence in support of the dosage-balance hypothesis has been lacking so far. Here, we use a stochastic birth-death model to investigate which gene families preferentially duplicate through WGM in the angiosperm lineage, and study their functional and evolutionary characteristics. Gene families that duplicate preferentially through WGM were found to exhibit stronger sequence divergence constraints and decreased rates of functional and expression divergence than gene families that show a volatile WGM/SSD duplication behavior, in accordance with the predictions of the dosage balance hypothesis. For at least six out of the top-10 reciprocally retained gene families recovered in our analyses, overexpression or deletion phenotypes reported in literature suggest that the function of the genes concerned is indeed dosage balance-sensitive.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

OM-PH5

Coordinated gene expression divergence following whole genome duplication in A. thaliana

Yves Van De Peer [1 2,*], Riet De Smet [1 2], Eshan Sabaghian [1 2], Zhen Li [1 2]

[1]Center for Plant Systems Biology, VIB, [2]Department of Plant Biotechnology and Bioinformatics, Ghent University, Zwijnaarde, Belgium

Abstract: It has been suggested that whole-genome duplication (WGD) plays an important role in the evolution of novel traits, yet little is known about the genetic basis underlying the evolution of phenotypic diversity associated with polyploidy. WGD copies entire genomes and as such creates the unique situation in which duplicated pathways could evolve novel functions by coordinated sub- or neofunctionalisation of its constituent genes. However, hitherto, few if any examples of such coordinated evolution of suites of genes following WGD are known for the flowering plants, where many WGD events have occurred. Here, we describe a set of 92 homoeologous gene pairs that show a similar pattern of tissue-specific gene expression divergence following WGD, with one homoeolog showing predominant expression in aerial tissues and the other homoeolog showing biased expression in tip-growth tissues (root-tip and pollen-tube). By including functional data, we provide evidence that this pattern of gene expression divergence involves genes functioning in the maintenance of cell-wall integrity and that following WGD these duplicated genes evolved separate functions in cell-growth and stress response.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

POA-156

Ancient polyploidy in orchids

Rolf Lohaus [1,*], Zhen Li [1], Guo-Qiang Zhang [2], Wen-Chieh Tsai [3 4 5], Yi-Bo Luo [6], Yves Van de Peer [1 7], Zhong-Jian Liu [2 8 9 10]

[1]Center for Plant Systems Biology, VIB, Ghent University, Gent, Belgium, [2]Shenzhen Key Laboratory for Orchid Conservation and Utilization, The National Orchid Conservation Center of China and The Orchid Conservation and Research Center of Shenzhen, Shenzhen, China, [3]Orchid Research and Development Center, [4]Department of Life Sciences, [5]Institute of Tropical Plant Sciences, National Cheng Kung University, Tainan, Taiwan, [6]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China, [7]Department of Genetics, Genomics Research Institute, Pretoria, South Africa, [8]The Center for Biotechnology and BioMedicine, Graduate School at Shenzhen, Tsinghua University, Shenzhen, [9]College of Forestry, South China Agricultural University, Guangzhou, [10]College of Landscape Architecture, Fujian Agriculture and Forestry University, Fuzhou, China

Abstract: With approximately 25,000 species, Orchidaceae is one of the two largest angiosperm families and represents a staggering 8–10% of the world's seed-plant species. Orchids are renowned for their spectacular flowers, with a huge diversity in floral designs and reproductive and ecological strategies, and have successfully colonized almost every habitat on Earth. Ever since Darwin's early fascination with orchids, they have hence attracted great interest from botanists and evolutionary biologists alike. Orchidaceae is divided into five subfamilies: Apostasioideae, Vanilloideae, Cypripedioideae, Orchidoideae and Epidendroideae. The genomes of the two orchids sequenced to date, *Dendrobium catenatum* and *Phalaenopsis equestris*, both from the Epidendroideae, contained signatures of ancient polyploidy events. Polyploidies, or whole genome duplications (WGDs), may have played a major role in the evolution of flowering plants and have been linked to diversifications and/or biological innovations. To improve our understanding of orchid genome evolution, we sequenced the genome of the early-diverging orchid *Apostasia shenzhenica* from Apostasioideae, the sister lineage to all other extant orchids. Here, we present phylogenomic analyses of all three orchid genomes combined with transcriptomes from the three other subfamilies of Orchidaceae to investigate the phylogenetic placement and timing of these ancient WGDs in orchid history.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

POA-145

The interplay of divergence and hybridization for speciation in evening primroses

Jesse Hollister [1,*]

[1]Department of Ecology and Evolution, Stony Brook University, Stony Brook, United States

Abstract: Hybridization is a frequent driver of speciation, particularly in plants. Here I examine the interaction of genomic divergence and hybridization in species formation in the plant genus *Oenothera* (evening primrose). This genus has repeatedly evolved a genetic system called Permanent Translocation Heterozygosity (PTH). The evolution of PTH involves chromosome translocations that break down recombination between unlike karyotypes, followed by establishment of structurally heterozygous individuals bearing both karyotypes. However, it is not known whether PTH lineages arose by gradual accumulation of translocations within a single progenitor species, or "all at once" by hybridization between different progenitor species with different karyotypes. Using population genomic analysis, I infer patterns of divergence and hybridization among *Oenothera* species, revealing a rich diversity of evolutionary histories. These inferences resolve debate regarding the processes that contributed to species-richness in this genus.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

POA-155

Patterns of genome-wide introgression in old world house mouse.

Eslam Elshahat [1,*], Michael Kohn

[1]BioSciences department, Rice University, Houston, United States

Abstract: Whole genome analyses are powerful tools to study the evolutionary history of species. Genomes of closely related species that hybridize can be used to study the role of introgression on the evolution of new traits, and to examine how genomes of biological species are porous to gene flow. The Western European house mouse (*Mus musculus domesticus*) hybridizes with the Algerian mouse (*Mus spretus*), a closely related species to the house mouse. We used PhyloNet–HMM, a recently developed Bayesian statistical method, to detect introgression in genomes of 4 introgressed mice. In addition, SNP arrays of 45 *M. m. domesticus* samples covering the geographical range of the species were scanned for introgression. These hybridization events left a genetic footprint of introgressed tracts across the house mouse genome ranging from 0.01% - 0.4%. Here, we further describe the genome wide distribution of introgressed tracts, and describe these in terms of their putative functional impact. We compared the full genome sequences of the house mouse and tested whether genes with non-synonymous mutations entered the populations of house mice as polymorphisms. We describe the spatial distribution and frequency of introgressed genes to infer selection on them. Lastly, we provide a detailed recombination landscape with regard to introgressed tracts.

Expanded summary*: Summary of the research – Recent evidence indicate the importance of genetic exchange between closely related species in genome evolution. However, this process is predicted to be mostly detrimental, and should be rare owing to ecological, behavioral and genetic isolation mechanisms. Introgressive hybridization, where species interbreed backcross results in polymorphism sharing and transferring, was recently reported between two mice species, *Mus musculus domesticus* and *Mus spretus*. The two species coexist in sympatry in North Africa and Spain, but introgression was found outside the area of sympatry. My dissertation is an attempt to understand the role of Introgressive hybridization in introducing adaptive mutations between closely related species, and to study the effect of introgression selection on species level.

High-density SNP genotyping arrays and whole genome sequences of *M. m. domesticus* and *M. spretus* were used to test for introgression between the two species. Hybridization has left a genome wide footprint, and introgressed genes appear of ecological and evolutionary relevance. For example, conferring resistance to warfarin rodent poison due to introgressed *Vkorc1* gene, and genes involved in e.g. olfaction and immunity. Hundreds of introgressed regions were found containing 1368 genes. I have investigated how far introgression reached outside the area of sympatry, the directionality of introgression, the amount of genetic material and genes exchanged. I am applying molecular, computational, and modeling tools throughout my dissertation.

Broad significance – The role of genome wide introgression as source of adaptive mutations is poorly understood. We managed to catch the process in the act between *Mus musculus domesticus* and *M. spretus*, where introgressed variants remain polymorphic in both species. My work will explores the types of selection acting upon introgression in natural populations (soft selective sweeps versus hard sweeps), to gauge whether adaptive introgression is rare phenomenon driven by strong selection, or more common and perhaps (initially) neutral. I aim to identify factors favoring the spread of introgressed genes outside the area of sympatry, and entering long term genomic variation of a species over, affecting the ecology and evolution of it.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

OM-PH4

The genetic and biochemical basis of floral color differences in Nicotiana allopolyploids

Elizabeth W. McCarthy [1,*], Andrea E. Berardi [2 3], Amber J. Lawhorn [1], Amelda Kurti [1], Stacey D. Smith [2], Amy Litt [1]

[1]Department of Botany and Plant Sciences, University of California, Riverside, Riverside, [2]Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, United States, [3]Institute of Plant Sciences, University of Bern, Bern, Switzerland

Abstract: Polyploidy is a widespread phenomenon in angiosperm evolution, and polyploids often display diverse or even transgressive (outside the range of the progenitors) phenotypes. Differences in floral phenotypes can lead to attraction of different pollinators, which can facilitate both reproductive isolation from progenitor species and species diversification following polyploidization. Here, we examine the genetic and biochemical basis of floral color diversity in *Nicotiana* polyploids.

Nicotiana (tobacco) is an excellent system in which to study polyploidy because about half of its species are allopolyploids (interspecific hybrids as well as polyploids) that range in age from first generation synthetic lines to 10 million year old species. *Nicotiana* species exhibit substantial floral diversity, and allopolyploids derived from the same progenitors often have distinct phenotypes. Here, we link floral pigment profiles to expression patterns of genes involved in the flavonoid biosynthetic pathway in accessions of the allotetraploid *Nicotiana tabacum*, as well as in first generation synthetic lines derived from the same progenitors, that differ in floral color. Magenta flowers have increased cyanidin concentration, but lower flavonol content than pale pink flowers. The pale pink phenotype is also correlated with dramatically increased expression of flavonoid biosynthetic pathway genes that produce flavonols relative to those that yield anthocyanins, suggesting that competition between these enzymes may affect floral color. In addition, natural *N. tabacum* accessions produce a small amount of transgressive pelargonidin, which is not generated in synthetic lines, suggesting that pelargonidin production likely evolved in the *N. tabacum* lineage, but not as a direct consequence of allopolyploidization.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

OM-PH6

Fitting two genomes in one nucleus: the structure and function of genomes in Epichloë hybrids

David Winter [1,*]

[1]Institute of Fundamental Sciences, Massey University, Palmerston Noth, New Zealand

Abstract: Allopolyploidy, the generation of new species from the merger of distinct parental lineages, has been an important evolutionary force in lineages spanning the eukaryotic tree of life. In addition to instantaneously generating new species, allopolyploidisation creates genetic novelty that can lead to adaptation and spark evolutionary radiations. For such novelty to flourish however, allopolyploids must first survive the shock that occurs when two divergent genomes occupy a single nucleus.
We have recently assembled the genome of fungus *Epichloë festucae* , a species which is a parent to many *Epichloë* allopolyploids. Combining this assembly with data from high throughput chromosome capture (HiC) and RNA seq experiments we show that the 3-dimensional structure of the genome within an *E. festucae* nucleus is partially determined by large blocks of repetitive DNA and that these structures influence gene expression. In particular, highly-expressed genes and those up-regulated when E. festucae is present in a plant tend to co-locate in the nucleus.

In this talk I will discuss progress on a new study examining how this structural regulation is maintained or altered in allopolyploids, where two sets of divergent chromosomes are present in a single nucleus. We are producing genome assemblies for *E. amarillans, E. elymi* and two allopolyploids that are the result of independent hybridisation between these species. Using HiC data we will then be able to determine if the 3-dimensional structure present in the parental species in maintained in allopolyploids, and how any alterations to this structure influence gene expression.

Expanded summary*: Allopolyploidy, the generation of new species from the merger of distinct parental lineages, has been an important evolutionary force in lineages spanning the eukaryotic tree of life. In addition to instantaneously generating new species, allopolyploidisation creates genetic novelty that can lead to adaptation and spark evolutionary radiations. For such novelty to flourish however, allopolyploids must first survive the shock that occurs when two divergent genomes occupy a single nucleus.
The fungal genus *Epichloë* provides a powerful model system in which to study the effects of allopolyploidy and reactions to genome shock. *Epichloë* species are endophytes that live in close association with grass species. Allopolyploid species dominate the genus, both in terms of the number of species (there are more than 30 hybrid species) and the prevalence of particular species in host populations. This genus also provides a number of practical benefits when compared to other polyploid systems. The allopolyploids are diploid (or rarely triploid) and the genomes are relatively small (30-50Mb), making high throughput genomic analyses relatively easy and inexpensive. Epichloë can also be cultured in the lab and be subjected to genetic and experimental manipulations.

We have recently assembled the genome of *E. festucae* , a parent to many of the allopolyploid species. Using PacBio sequencing we have been able to generate a complete chromosomal assembly, including full-length sequences of all repeats. Combining this assembly with data from high throughput chromosome capture (HiC) and RNA seq experiments we show that the 3-dimensional structure of the genome within an *E. festucae* nucleus is partially determined by large blocks of repetitive DNA and that these structures influence gene expression. In particular, highly-expressed genes and those up-regulated when E. festucae is present in a plant tend to co-locate in the nucleus.

In this talk I will discuss progress on a new study examining how this structural regulation is maintained or altered in allopolyploids, where two sets of divergent chromosomes are present in a single nucleus. We are producing PacBio assemblies for *E. amarillans, E. elymi* and two allopolyploids that are the result of independent hybridisation between these species. Using HiC data we will then be able to determine if the 3-dimensional structure present in the parental species in maintained in allopolyploids, and how any alterations to this structure influence gene expression. In addition to examining the structure of the hybrid genomes, this data will allow us to detect recombination among parental genomes and measure differences in transposon activity and gene expression between parental

and hybrid species. Taken together, these results will allow us to measure the structural and functional responses to genome shock following two independent polyploidization events.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

OM-PH1

Regulatory and functional divergence of duplicated genes after ancient polyploidy events

Shao-Lun Liu, Yichun Qiu, Andrej Arsovski, Keith Adams*

Abstract: Numerous episodes of paleopolyploidy have occurred during flowering plant evolution and as a result, many plants have several polyploidy events in the evolutionary history of their lineage. Duplicated genes can have a variety of fates. I will present projects on changes in subcellular localization, concerted divergence of gene products that function together, and divergence in cis-regulatory elements and regulatory networks. We analyzed experimental localization data from green fluorescent protein experiments for 128 duplicate pairs in Arabidopsis thaliana, revealing 19 pairs with subcellular relocalization. Many more of the duplicate pairs with relocalization than with the same localization showed an accelerated rate of amino acid sequence evolution in one duplicate, and one gene showed evidence for positive selection. We identified potential sequence mutations through comparative analysis that likely result in relocalization of two duplicated gene products. We show that four cases of relocalization have new expression patterns, compared with orthologs in outgroup species, including two with novel expression in pollen. Next we provide an example of concerted divergence of simultaneously duplicated genes whose products function in the same complex: the PRC2 complexes FIS2 and VRN2 in Brassicaceae. The VRN-PRC2 complex, which functions in vernalization and the control of flowering time, contains VRN2 and SWN, and both genes were duplicated by polyploidy to generate FIS2 and MEA which function in the Brassicaceae-specific FIS-PRC2 complex that regulates seed development. We present multiple lines of evidence indicating that *FIS2* and *MEA* have diverged in concert, resulting in functional divergence of the PRC2-complexes in Brassicaceae. Finally we analyzed a comprehensive data set of DNase I sequencing-identified cis-regulatory binding sites at single-base-pair resolution to compare binding sites and network connectivity in duplicated gene pairs. We found that duplicated gene pairs vary greatly in their cis-regulatory element architecture, resulting in changes in regulatory network connectivity. Many of the duplicated genes show partial or complete divergence in their network connections. These studies illustrate various ways in which duplicate gene divergence occurs after ancient polyploidy events.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

POA-149

Elucidating the Multiple Independent Whole Genome Duplication Events in the Core Brassicales

Makenzie Mabry*, Julia Brose [1], Wade Dismukes [2], Aleksandra Beric [3], Jacob Washburn [1], Pat Edger [4], Jocelyn Hall [5], Michael McKain [3], Ihsan Al-Shehbaz [6], Alex Harkess [3], M. Eric Schranz [7], Gavin Conant [8], J. Chris Pires [1]

[1]University of Missouri, Columbia, [2]Iowa State University, Ames, [3]Donald Danforth Plant Center, St. Louis, [4]Michigan State University, East Lansing, United States, [5]University of Alberta, Edmonton, Canada, [6]Missouri Botanical Garden, St. Louis, United States, [7]Wageningen University and Research, Wageningen , Netherlands, [8]North Carolina State University, Raleigh, United States

Abstract: The Brassicales are an economically important order of flowering plants. Many crop species such as Kale, Broccoli, Cabbage, Cauliflower, Canola oil, Capers, and Papaya as well as the model plant organism, *Arabidopsis thaliana*, all belong to this diverse order. Until now, studies have either used a few genes or a few taxa to understand the relationships within this group of plants. Phylo-Transcriptomics, a quickly evolving field that allows for the use of RNA-seq data to be used to make phylogenomic inferences by allowing access to many more nuclear genes than using traditional PCR is a great method to no longer have to choose between too few genes or taxa. Using this new method, both the relationships and evolution of independent gene and genome duplication events is assessed to help understand this rich diversity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

OM-PH3

Genomic variations in the nascent generations of an allopolyploidy line by goldfish and common carp

Jing Chai [1,*], Li Ren [2 3], Min Tao [2 3], Shaojun Liu [2 3], Yaping Zhang [1 4], Jing Luo [1]

[1]State Key Laboratory for Conservation and Utilization in Yunnan, and School of Life Sciences, Yunnan University, Kunming, [2]College of Life Sciences, [3]State Key Laboratory of Developmental Biology of Freshwater Fish, Hunan Normal University, Changsha, [4]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, the Chinese Academy of Sciences, Kunming, China

Abstract: Polyploidization or genome duplication triggers vast genome instability including random gene loss, accelerated mutations, chromosomal rearrangements and failed paring of homologous chromosomes, and these synthetic effects are defined as genome shock. The outcome of combining two genomes in vertebrates remains unpredictable especially because polyploidization seldom shows positive effects, and more often results in lethal consequences because viable gametes fail to form during meiosis. Fortunately, goldfish (maternal♀) × the common carp (paternal♂) hybrids have reproduced successfully up to generation 26 and this permits an investigation into the genomics of hybridization and polyploidization. The first two generations of these hybrids are diploids and subsequent generations are tetraploids. Based on the reference genome of goldfish, the analyses of resequencing data from parents and six generations (20 samples in total) were performed. The results showed that, (1) variations distributed in untranslated regions (UTR) were significantly more abundant than the ones in coding regions (CDS) and intron; the six individuals of first generation $2nF_1$ were with the difference of possessing variations; (2) Insertions/Deletions (InDels) distributed in UTR regions were with significantly higher ratios than in intron and CDS; (3) within 13 offspring, chimeric genes comprised about 9.75%–11.21%; six individuals of $2nF_1$ shared only 2,559 chimeric genes, which indicated the individual difference of variation pattern within this generation; the chimeric genes with mutations were enriched in the regulation of physiological functions and immunoreactivity, and the regulation of programmed cell death and apoptosis. The discoveries yielded by resequencing data indicated fast changes including the genomic variations, especially in $2nF_1$. These syndrome effects caused by genome shock might be the reasons of the greatly reduced viability in $2nF_2$ hybrid offspring, which might be lethiferous in natural polyploids. The results might help explain why polyploid are much rarer in vertebrates than in plants. Ultimately, further work remains on exactly how polyploid plants and animals survive from genome shock, which happens more frequently in allopolyploid plants than in animals. Additional functional analyses at the genomic (genetic/epigenetic) level are also awaited. In the future work, we expect to find out possible key regions in the genomes and key changes in the gene pathways and regulatory elements which are most importantly contributed to the survival and genomic stability for the allopolyploidy.

Keywords: Allopolyploidization; Vertebrates; Chimeric genes; Re-sequencing

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

POA-151

Is there a parent-of-origin effect on sex-biased gene expression in a hybrid plant?

Melissa Toups [1,*], Guillaume Cossard [2], John Pannell [2]

[1]Integrative Biology, University of Texas, Austin, United States, [2]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

Abstract: Virtually nothing is known about how sexually dimorphic gene expression evolves in an allopolyploid with separate sexes. We asked whether the sex of a homeolog's parent-of-origin affects its sex-biased expression in the hybrid species, *Mercurialis annua*. Populations of *M. annua* are comprised of males and hermaphrodites. We conducted RNA-seq on pools of males, of hermaphrodites, and of each of its progenitor species. We ask whether male-biased genes are more likely to be derived from the male parent, and if hermaphrodite-biased genes are more likely to be derived from the hermaphroditic parent.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

POA-150

Bayesian inference of phylogenetic networks using sequence data from multiple unlinked loci

Luay Nakhleh*, Dingqiao Wen

Abstract: Using genome-wide data for phylogenetic inference and analysis has become very common in the post-genomic era, giving rise to the field of *phylogenomics*. The multispecies coalescent (MSC) model has emerged as the main stochastic process that helps capture the intricate relationship between species trees and gene trees. Combined with models of sequence evolution, the MSC can be viewed as a generative model of genomic sequence data in the context of a (species) phylogenetic tree. In particular, the MSC naturally explains and allows for quantifying the phenomenon of incomplete lineage sorting (ILS).
From a biological perspective, a significant outcome of the use of genome-wide data has been the increasing evidence, or hypotheses, of reticulation during the evolution of various groups of eukaryotic. Therefore, developing models and methods that allow for reticulation in evolutionary analyses is very important. In this talk, I will describe the multispecies network coalescent (MSNC) model, which extends the MSC model so that it operates within the branches of a phylogenetic network. This extended model naturally allows for modeling vertical and horizontal evolutionary processes acting within and across species boundaries. In particular, it simultaneously accounts for gene tree incongruence across loci due to both hybridization and incomplete lineage sorting. I will then describe a likelihood function for this model, as well as a method for Bayesian sampling of phylogenetic networks and their parameters. The method uses as data sequence alignments of multiple unlinked loci. I will demonstrate its use via the PhyloNet software package.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

OM-PH2

Polyploidy, Paths and Pitfalls to a Paradigm: Insights from Tragopogon

Douglas Soltis*, Pamela Soltis [1]

[1]Florida Museum of Natural History, University of Florida, Gainesville, United States

Abstract: Polyploidy (whole-genome duplication; WGD) is a major force in eukaryotic evolution, and ancient and recent WGDs are well documented in many lineages, including vertebrates, fungi, and, most extensively, angiosperms. Much of what we know about polyploidy comes from studies of older events in crops, or the immediate impact in synthetic lines. Few systems provide a range of ages for examining the process of polyploidy and little is known about the earliest stages of polyploidy and the processes that transform a duplicated genome into a diploid one. *Tragopogon* (Compositae) s a textbook model for studying polyploidy with two recently (~80 years old; 40 generations in these biennials) and repeatedly formed natural allotetraploids (*T. mirus*, *T. miscellus*) and their diploid parents (*T. dubius*, *T. pratensis*, *T. porrifolius*). Our cytogenetic, genetic and genomic data suggest that patterns of gene loss, silencing, and chromosomal change in the young polyploids *T. mirus*, *T. miscellus* are repeated across multiple origins of these polyploids in nature, a well as in synthetic lines. These changes are ongoing--we have caught evolution in the act. Furthermore, the ongoing changes observed in these newly formed polyploids appear to be fixed in the older polyploid *T. castellanus*. Lastly, some of the changes detected mirror changes that occurred deep in time following WGD near the origin of the Compositae. Thus, patterns of genetic and genomic change may be predictable within just a few generations in young polyploids. With its range of ages, *Tragopogon* affords research opportunities not available in other systems (e.g., cotton, soybean, maize).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

POA-152

Tissue specific ploidy variation in sexual and apomictic seeds

Dorota Paczesniak*, Marco Pellino [1], John T. Lovell [2], Devan Guenter [1], Siegfried Jahnke [3], Johanna Roussel [3 4], Andreas Fischbach [3], Timothy F. Sharbel [1]

[1]Global Institute For Food Security, Saskatoon, Canada, [2]Department of Integrative Biology, The University of Texas, Austin, United States, [3]IBG-2: Plant Sciences, Jülich Forschungszentrum, [4]FH Aachen, University of Applied Sciences, Jülich, Germany

Abstract: Polyploidy effects are often examined at the level of a whole organism or across species, however ploidy variation occurs also at intra-individual level and can be limited to a specific tissue, e.g. endosperm tissue in the seeds of flowering plants.
Sexual reproduction typically involves double fertilization, whereby the egg and central cells of the ovary are fertilized to produce a diploid embryo and a triploid endosperm (nutritious tissue with 2 maternal and 1 paternal genomes). In the genus *Boechera*, a wild relative of *Arabidopsis*, both sexual and apomictic (i.e. reproducing asexually via seed) lineages are found. Diploid apomictic lineages produce a meiotically-unreduced diploid egg cell which develops parthenogenetically (without fertilization) into an embryo that is genetically identical to the mother plant, whereas endosperm is pseudogamous (i.e. requires fertilization by diploid pollen) and typically hexaploid, with 2:1 maternal to paternal genome ratio. Hence the seeds of sexual and apomictic Boechera have the same embryo ploidy (2x), but differ with respect to endosperm ploidy (3x and 6x).
We are interested in understanding effects of ploidy variation on endosperm function and plant fitness, in the context of sexual and apomictic reproduction. To do so we have (1) compared gene expression patterns between developing hexaploid (apomictic) and triploid (sexual) endosperm in the seeds from multiple apomictic and sexual genotypes using a custom *Boechera* microarray and live microdissected endosperm tissue, and (2) have examined the effects of genotype, ploidy and parent of origin effects as potential factors underlying phenotypic variation in seed size, an agriculturally important trait.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

POA-144

Whole genome duplications, self-incompatibility evolution and lineage-specific speciation rates are strongly associated within Brassicaceae

Xavier Vekemans [1,*], Laura Henocq [1], Sophie Gallina [1], Vincent Castric [1], Celine Poux [1]

[1]UMR Evo-Eco-Paleo, University of Lille, Lille, France

Abstract: Recent studies using phylogenetic approaches, mapping of trait evolution, and quantification of species diversification have shown that transitions from outcrossing to selfing and from diploidy to polyploidy are associated with large reductions in species diversification rates. However, the joint evolution of polyploidy and mating system genes has rarely been addressed. We studied the impact of whole genome duplication events on the evolutionary dynamics of the self-incompatibility (SI) system in Brassicaceae. Loss of SI and shift towards selfing is a common feature of neopolyploid lineages, and thus the observed reduction in rates of species diversification in those lineages could be mostly due to its indirect effect on mating system change. However, retrospective studies indicate that a few independent mesopolyploid lineages seem to have retained an active SI system and thus an outcrossing mating system. In three of those lineages we found a clear signature of strong genetic bottleneck, followed by allelic re-diversification at the S-locus. In all three cases, this bottleneck appears to be associated with the historical event of polyploidy. Detailed analyses of the S-locus genomic region in two cases have revealed common patterns (deletion of the S-locus in its ancestral position, evolution of a new S-locus at a different genomic position), as well as striking differences (the genes involved in pollen-pistil recognition at the new S-locus are either orthologous or non-orthologous to the *SCR* and *SRK* genes of the ancestral Brassicaceae). These results suggest a scenario with a temporary loss of self-incompatibility after polyploidization in those lineages, followed by re-establishment of SI at a different genomic location with rapid allelic re-diversification caused by strong balancing selection. Although in most cases polyploidy is accompanied by loss of SI, selfing evolution and reduction of species diversification rates, we suggest that the re-establishment of SI in a few lineages after polyploidization has been the trigger for lineage-specific species diversification and long-term evolutionary success.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Polyploidy and hybridization

POA-154

Evidence for congruent three-genome phylogenetic signals for four botanical sections of the flax genus Linum

Yong-Bi Fu [1,*]

[1]Plant Gene Resources of Canada, Agriculture and Agri-Food Canada Saskatoon RDC, Saskatoon, Canada

Abstract: Previous studies of phylogenetic relationships of plant species using DNA sequences from chloroplast, mitochondrial and nuclear genomes frequently reported incongruent phylogenetic signals. Our presentation will reveal the evidence for congruent phylogenetic signals acquired from chloroplast, mitochondrial and nuclear genomes for four botanical sections of the flax genus *Linum*. It was obtained through the development of a multiplexed shotgun sequencing (MSS) protocol and 3GenomeSNP pipeline to assess the phylogenetic relationships of 18 *Linum* samples representing 16 species within four botanical sections of the genus. Also, we will present the MSS protocol and 3GenomeSNP pipeline and discuss their utility in acquisition of three-genome phylogenetic signals of non-model organisms

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-131

Reconstructing the evolutionary history of present-day and extinct elephants

Eleftheria Palkopoulou [1,*], Mark Lipson [1], Swapan Mallick [1], Jacob Enk [2], Emil Karpinski [3], Nadin Rohland [4], Ross Macphee [5], Grant Zazula [6], Matthias Meyer [7], Kurt Alt [8], Harald Meller [9], Stefan Claesson [10], Elephant Genome Sequencing Consortium [11], Hendrik Poinar [12], Michael Hofreiter [13], David Reich [1]

[1]Department of Genetics, Harvard Medical School, Boston, United States, [2]Departments of Anthropology and Biology, and the Michael G. DeGroote Institute for Infectious Disease Research, [3]McMaster University, Hamilton, Canada, [4]Harvard Medical School, Boston, [5]Division of Vertebrate Zoology/Mammalogy, American Museum of Natural History, New York, United States, [6]Department of Tourism and Culture, Yukon Government, Whitehorse, Canada, [7]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, [8]Institute for Anthropology, Johannes Gutenberg-University Mainz, Mainz, [9]Landesamtes für Denkmalpflege und Archäologie Sachsen-Anhalt, Halle, Germany, [10]Search, Portsmouth, [11]Broad Institute, Boston, [12]Departments of Anthropology and Biology, and the Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, United States, [13]Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany

Abstract: African elephants (*Loxodonta sp.*) and Asian elephants (*Elephas maximus*) are the only extant lineages of the family *Elephantidae* that emerged in Africa during the late Miocene about 9 million years ago. From their extinct relatives, woolly mammoths (*Mammuthus primigenius*) disappeared just a few thousand years ago, while others, such as the straight-tusked elephants (*Paleoloxodon antiquus*) are believed to have become extinct earlier by the end of the previous interglacial period. Genetic studies have established parts of the proboscidean phylogeny, showing that African forest (*L. cyclotis*) and savanna (*L. africana*) elephants comprise different species as well as that extinct woolly mammoths were most closely related to Asian elephants. However, other taxa, such as the European straight-tusked elephant, have not been genetically analyzed to date and their phylogenetic relationship to other elephantid lineages are solely based on morphological data. We generated high-coverage genomes from 2 forest elephants, 2 savanna elephants, 2 Asian elephants and 1 straight-tusked elephant, as well as lower coverage genome-wide data from 4 woolly mammoths, 1 Columbian mammoth, 1 straight-tusked elephant and 2 mastodons (*Mammut americanum*, a proboscidean outgroup). We analyzed this dataset to document the phylogenetic relationships among elephantid lineages, and reconstruct gene flow, demographic changes and natural selection. The evolutionary history of elephants appears to have been rather complex with multiple admixture events between different elephantid taxa, such as between woolly and Columbian mammoths, and between forest and straight-tusked elephants.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

OW-POG6

Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods

Johannes Krause*, Verena Schuenemann, Alexander Peltzer, Beatrix Welte, W. Paul van Pelt, Martyna Molak, Chuanchao Wang, Anja Furtwaengler, Chirstian Urban, Ella Reiter, Kay Nieselt, Barbara Tessmann, Michael Francken, Katerina Harvati, Wolfgang Haak, Stephan Schiffels

Abstract: Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman period

Johannes Krause[1,2,3], Verena J. Schuenemann[2,3], Alexander Peltzer[1,4], Beatrix Welte[2], W. Paul van Pelt[5], Martyna Molak[6], Chuan-Chao Wang[1], Anja Furtwängler[2], Christian Urban[2], Ella Reiter[2], Kay Nieselt[4], Barbara Teßmann[7], Michael Francken[2], Katerina Harvati[2,3,8], Wolfgang Haak[1], Stephan Schiffels[1]

1. Department for Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany.
2. Institute for Archaeological Sciences, University of Tübingen, 72070 Tübingen, Germany.
3. Senckenberg Center for Human Evolution and Paleoenvironment, 72070 Tübingen, Germany.
4. Integrative Transcriptomics, Center for Bioinformatics, 72070 Tübingen, Germany.
5. Division of Archaeology, University of Cambridge, Cambridge CB2 3DZ, United Kingdom
6. Museum and Institute of Zoology, Polish Academy of Sciences, 00-679 Warsaw, Poland.
7. Berlin Society of Anthropology, Ethnology and Prehistory, 10997 Berlin, Germany.

Abstract:

Egypt, located on the isthmus of Africa, is an ideal region to study historical population dynamics due to its geographic location and documented interactions with ancient civilizations in Africa, Asia, and Europe. Particularly, in the first millennium BCE Egypt endured foreign domination leading to growing numbers of foreigners living within its borders possibly contributing genetically to the local population. Here we present 90 mitochondrial genomes as well as genome-wide datasets from three individuals obtained from Egyptian mummies. The samples recovered from Middle Egypt span around 1,300 years of ancient Egyptian history from the New Kingdom to the Roman Period. Our analyses reveal that ancient Egyptians shared more ancestry with Near Easterners than present-day Egyptians, who received additional Sub-Saharan admixture in more recent times. This analysis establishes ancient Egyptian mummies as a genetic source to study ancient human history and offers the perspective of deciphering Egypt's past at a genome-wide level.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-130

Short term reduction in Pan troglodytes schweinfurthii genetic diversity at Gombe National Park

Andrew Ozga [1,*], Maria Nieves-Colon [2], Timothy Webster [3], Melissa Wilson Sayres [3 4], Rebecca Nockerts [5], Michael Wilson [5 6], Ian Gilby [1 2], Anne Pusey [7], Anne Stone [1 2 4]

[1]Institute of Human Origins, [2]School of Human Evolution and Social Change, [3]School of Life Sciences, [4]Center for Evolution and Medicine, Arizona State University, Tempe, [5]Department of Anthropology, [6]Department of Ecology, Evolution, and Behavior, University of Minnesota, Minneapolis, [7]Evolutionary Anthropology Department, Duke University, Durham, United States

Abstract: Human induced habitat loss due to agricultural expansion, wild game hunting, and zoonotic disease transmission are all pressing concerns for the wellbeing of wild non-human primates. In particular, habitat destruction and deforestation, which has been occurring at Gombe National Park in Tanzania since the 1990s, directly impact resource availability and habitat connectivity which in turn are expected to influence a species' genetic diversity. However, the extent to which this has impacted chimpanzee (*Pan troglodytes schweinfurthii*) genetic diversity has not been fully explored. Using three different whole genome and exome capture kits along with Illumina sequencing, we generated genomic data from 4 living chimpanzees (using feces) and 2 deceased chimpanzees (using dental calculus and dentine) from Gombe National Park. We combined these data with publicly available data and complete mitochondrial genomes (98-100% coverage) from additional individuals (9 dentine samples and 28 calculus samples) to characterize diversity within this potential genomic 'island.' Specifically, we found slightly higher levels of mitochondrial HV1 haplotype diversity in chimpanzees who died before 1992 than those who died after (0.862 compared to 0.824) suggesting a decrease in overall diversity occurring over a relatively short period of time. Additionally, this study is the first to reconstruct host genomic data from deceased non-human primate calculus.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

OW-POG4

Ancient mitogenomes of Argentine Patagonia (6070-310 YBP) reveal the early contribution of lineages not previously found in South America

Maria Laura Parolin [1,*], Rosa Irene Fregel Lorenzo [2], Claudio Marcelo Bravi [3], Fernando Luis Mendez [2], Silvia Lucrecia Dahinten [1], Julieta Gómez Otero [1], Beth Shapiro [4], Richard Edward Green [4], Camila Tamburrini [1], Néstor Guillermo Basso [1], Carlos Bustamante [2]

[1]Instituto de Diversidad y Evolución Austral, CENPAT-CONICET, Puerto Madryn, Argentina, [2]Department of Genetics, Stanford University, Stanford, United States, [3]Instituto Multidisciplinario de Biología Molecular , CONICET CICPBA: UNLP, La Plata, Argentina, [4]Paleogenomics Lab, University of California Santa Cruz, Santa Cruz, United States

Abstract: Patagonia, a *cul-de-sac* for the continental dispersal of *Homo sapiens* into the Americas, has been extensively analyzed by archaeologists and bioanthropologists. Despite its deep and rich archaeological record, the genetic make-up of its pre-contact inhabitants is barely known. We present the analysis of 19 mitogenomes recovered by NGS from archaeological sites in Central Patagonia (Chubut, Argentina). 15 are pre-contact samples, (6,010-770YBP), and four are samples from historical times (550-310YBP). Preservation of the samples was extraordinary, with endogenous DNA values up to 68%. Preliminary analysis indicates that mtDNA genomes from Central Patagonia belong to haplogroups B2 (21%), C1b (21%), C1c (5%), D1 (47%) and D4h3a (5%). This is surprising when compared with a database of >19,000 control region sequences and >2,100 mitogenomes of Native American origin: 1) While modern indigenous populations from Northern and Central Patagonia of Argentina and Chile derive 41-54% of their maternal lineages from clades B2i2 and C1b13, none of these are present in our dataset; 2) 75% of our B2 and C1b lineages share derived polymorphism with lineages so far known to be present only in modern Central-Western Argentina; and 3) although modern Patagonians carry the highest continental frequencies of D1g, the ancient set is enriched mostly in D1g5, a lineage widely distributed from South-Central Andes to Tierra del Fuego (including two samples with private motifs not described before). The upcoming analysis of the nuclear portion of these samples will help us better understand migratory and admixture processes in the Patagonian region.

Expanded summary*: As mentioned in the abstract, this research constitutes the first analysis of complete ancient mtDNA genomes of Argentine Patagonia (6070-310 YBP) revealing the early contribution of lineages not previously found in South America, and it will represent a great opportunity to discuss it with peers in the SMBE meeting.

The Austral Evolution and Diversity Institute (*Instituto de Diversidad y Evolución Austral* -IDEAus) in Centro Nacional Patagonico (CENPAT-CONICET), where I have been a researcher since 2012, is located in Patagonia Argentina. During the last 10 years, I have been leading and collaborating in different bio-anthropological research projects whose main goal is to increase our understanding of the "genetic identity" of the Argentinean population, trying to demystify the widespread belief that the country may have the highest European ancestry in South America. On the contrary, several anthropo-genetic studies from different regions of the country demonstrated that the native component increases towards the north and south of the country while the European component is concentrated in the center of Argentina.

During the last 7 years, I have been devoted almost exclusively to the anthropo-genetic and demographic study of urban and rural not previously studied populations from Patagonia. My interest arises not only from the fact that I am native of Patagonia, but also because this region is one of the least studied places in terms of its genetic characterization. Focusing on singularities as a result of different migratory histories and differential admixture processes in each locality. In collaboration with archeologists, anthropologists, and geneticists from the IDEAus, we recently started a new research project to use genetics in study of past Patagonian populations. Anthropologists from IDEAus have worked for more than 25 years in the recovery and study of human remains and artifacts associated with the lifestyle of hunter-gatherer populations from Patagonia (called Tehuelche), which inhabited the centro-coastal region. As mentioned in the abstract, both, the history of human settlements in South America and the biological diversity of the hunter-gatherer populations inhabiting the southern tip of the continent, have been extensively studied by researcher from multiple

disciplines. Recently, two archaeological sites that evidence early human settlements in Chile and Argentina (14,500 BP) have been reported. However, there is to date scarce genetic information available from the early human populations of Argentina, and in particular, from the Patagonia region.

Thanks to a fellowship granted by CONICET, to the support of Dr. Carlos Bustamante and his team (Department of Genetics, Stanford University), and to that of the Paleogenomics Lab at the University of California Santa Cruz (for the processing of the archaeological samples), I have been able to retrieve information on new mitochondrial lineages. While we have sequenced 19 individuals in this work. IDEAus holds an osteological collection with more than 120 samples from individual and common burials, all of which are archaeologically contextualized with dates ranging from the Middle to the Late Holocene and from historical times. The upcoming analysis of this large collection (the largest in Patagonia) will contribute to a broader understanding of evolutionary and population processes involving *Homo sapiens* populations as they migrated to the southern end of the Americas.

Given the novelty of our results and its relevance to some of the major topics in the 2017 SMBE meeting, attending this event would constitute a great personal opportunity to present the result of our aDNA work in an international meeting. Likewise, the annual SMBE meeting is particularly important for young researchers, especially for Latin America, because it allows us to present our work to an expert audience and provides us with the opportunity to enlarge our networks and create fruitful connections, like the one that made this project possible.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

OW-POG2

Empirical perspectives on ancient DNA

Maanasa Raghavan [1,*]

[1]Department of Zoology, University Of Cambridge, Cambridge, United Kingdom

Abstract: Technological advances in sequencing have propelled ancient DNA research into the paleogenomics era, accompanied by a tremendous increase in the amount of generated data. Additionally, continued efforts towards optimizing laboratory pipelines, including sampling, DNA extraction and library preparation techniques, have resulted in increased DNA yield from ancient samples. These genome-wide datasets are being used to reconstruct past population structures and understand how processes such as ancient migrations, admixture and selection have shaped present-day gene pools. My talk will focus on the contributions of ancient DNA research, in tandem with archaeology and anthropology, to the emerging picture of modern human dispersals and provide perspectives on how recent collaborative efforts with fields such as ecology and pathology are providing a new dimension to our understanding of the human past.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-128

Including sequencing error and alignment uncertainty into coalescence-based analyses

Peter Beerli [1,*], Kyle Shaw [1]

[1]Scientific Computing, Florida State University, Tallahassee, United States

Abstract: Population genetics analyses depend on the variable sites in the dataset. Modern sequencing methods are not perfect and include sequencing errors. Commonly, sequences are reported with quality scores. Often, these quality scores are discarded by analysis methods for population genetics models. We developed a method that can take aligned data and these quality scores as input to our program MIGRATE 4.3. These quality scores inform the likelihood calculation of the sequence uncertainty. We also use the Poisson-indel process to take into account gaps in the alignment. Taking into account sequencing quality scores and alignment uncertainties allow improving the estimation of parameters, such as population size, migration, and divergence.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-150

Western Europe during the third millennium BCE: A genetic characterization of the Bell Beaker Complex

Iñigo Olalde [1,*], Nadin Rohland [1], Swapan Mallick [1 2 3], Nick Patterson [2], Iosif Lazaridis [1 2], Alissa Mittnik [4 5], Selina Brace [6], Anna Szecsenyi-Nagy [7 8], Nasreen Broomandkhoshbacht [1], Matthew Ferry [1], Eadaoin Harney [1], Megan Michel [1], Jonas Oppenheimer [1], Kristin Stewardson [1], Eveline Altena [9], Peter De Knijff [9], Alistair Barclay [10], Jacqueline I. Mckinley [10], Eszter Banffy [11], David Billoin [12], Thomas Booth [13], Oliver Craig [14], Gordon Cook [15], Yoan Diekmann [16], Zuzana Faltyskova [16], Anthony Denaire [17], Michal Ernée [18], Milan Kuchařík [19], Joan Francès Farré [20], Harry Fokkens [21], Michiel Gazenbeek [22], Volker Heyd [23], Kristian Kristiansen [24], Philippe Lefranc [25], Olivier Lemercier [26], Arnaud Lefebvre [27], Tona Majó [28], Pierre-Jérôme Rey [29], Joël Serralongue [30], Philipp W. Stockhammer [31], Luc Vergnaud [32], João Zilhão [33 34 35], Kurt W. Alt [8 36 37], Mark G. Thomas [38], David Caramelli [39], Ron Pinhasi [40], Ian Barnes [41], Eske Willerslev [42], Johannes Krause [43], Morten Allentoft [42], Wolfgang Haak [43 44], Ian Armit [45], Carles Lalueza-Fox [46], David Reich [1 2 3]

[1]Department of Genetics, Harvard Medical School, Boston, [2]Broad Institute of MIT and Harvard, Cambridge, [3]Howard Hughes Medical Institute, Harvard Medical School, Boston, United States, [4]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, [5]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany, [6]Department of Earth Sciences, Natural History Museum, London, United Kingdom, [7]Laboratory of Archaeogenetics, Institute of Archaeology, Research Centre for the Humanities, Budapest, Hungary, [8]Institute of Anthropology, Johannes Gutenberg University of Mainz, Mainz, Germany, [9]Dept. of Human Genetics, Leiden University Medical Center, Leiden, Netherlands, [10]Wessex Archaeology, Salisbury, United Kingdom, [11]Archaeological Institute of the Hungarian Academy of Sciences, Budapest, Hungary, [12]INRAP, Institut National de Recherches Archéologiques Préventives, BUFFARD, France, [13]Natural History Museum, London, [14]Department of Archaeology, University of York, York, [15]University of Glasgow, Glasgow, [16]Research Department of Genetics, Evolution and Environment, University College London, London, United Kingdom, [17]Université de Strasbourg, Strasbourg, France, [18]Academy of Sciences of the Czech Republic, Institute of Archaeology, [19]Labrys o.p.s., Prague, Czech Republic, [20]Museu i Poblat Ibèric de Ca n'Oliver, Cerdanyola, Spain, [21]Leiden University, Leiden, Netherlands, [22]INRAP, Institut National de Recherches Archéologiques Préventives, Nice, France, [23]University of Bristol, Bristol, United Kingdom, [24] University of Gothenburg, Gothenburg, Sweden, [25]INRAP, Institut National de Recherches Archéologiques Préventives, Strasbourg, [26] Université Paul Valéry, Montpellier, [27]INRAP, Institut National de Recherches Archéologiques Préventives, Bordeaux, France, [28]Archaeom. Departament de Prehistòria, Universitat Autònoma de Barcelona, Barcelona, Spain, [29]Université Savoie Mont Blanc, [30]Service archéologique, Conseil Général de la Haute-Savoie, Chambéry, France, [31]Institut für Vor- und Frühgeschichtliche Archäologie und Provinzialrömische Archäologie, Ludwig-Maximilians-Universität München, Munich, Germany, [32]ANTEA Bureau d'étude en Archéologie, Habsheim, France, [33]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain, [34]Centro de Arqueologia, Universidade de Lisboa, Lisboa, Portugal,

[35]Dept. Prehistòria, H. Antiga i Arqueologia, Universitat de Barcelona, Barcelona, Spain, [36]Center of Natural and Cultural History of Teeth, Danube Private University, Krems, Austria, [37]Institute for Integrative Prehistory and Archaeological Science, University of Basel, Basel, Switzerland, [38]Department of Genetics, Evolution, and Environment, University College London, London, United Kingdom, [39]University of Florence, Florence, Italy, [40]Department of Anthropology, University of Vienna, Vienna, Austria, [41]Department of Earth Sciences, Natural HistoryMuseum, London, United Kingdom, [42]Centre for GeoGenetics, Natural History Museum, University of Copenhagen, Copenhagen, Denmark, [43]Max Planck Institute for the Science of Human History, Jena, Germany, [44]Australian Centre for Ancient DNA, School of Biological Sciences, University of Adelaide, Adelaide, Australia, [45] University of Bradford, Bradford, United Kingdom, [46]Institute of Evolutionary Biology , CSIC-Universitat Pompeu Fabra, Barcelona, Spain

Abstract: The Bell Beaker Complex (BBC) was a widely scattered western European archaeological phenomenon that probably first appeared in Late Neolithic Iberia around 2,800 BCE, before spreading north and east and finally disappearing ~1,800 BCE. An open question is the extent to which the cultural elements associated with the BBC spread through movement of ideas or people. We present new genome-wide DNA data from 196 Neolithic and Bronze Age Europeans – the largest report of genome-wide data in a single study to date – and merge it with published data to form a dataset with 109 BBC individuals that provides a genomic characterization of the BBC across its geographic and temporal range. In contrast to people of the Corded Ware Complex (CWC) who were partly contemporaries of the BBC in central and eastern Europe and who brought steppe ancestry into central Europe through mass migration and replacement of local populations, our data suggest that the initial spread of the BBC into central Europe from the Iberian Peninsula was not mediated by a large-scale migration but rather through communication of ideas. However, we also show that the further spread of the BBC beyond central Europe did involve substantial movement of people. Focusing on Britain, which includes 81 of our new samples in a time transect from 3,900-1,300 BCE, we show that the arrival of the BBC ~2,400 BCE was mediated by migration from the continent: British individuals associated with Beakers are genetically indistinguishable from continental individuals associated with the same material culture and genetically nearly completely discontinuous with the previously resident population. This discontinuity persists through to samples from the Bronze Age – documenting a demographic turnover at the onset of the Bronze Age that is crucial in understanding the formation of the present-day British gene pool. The arrival of the BBC in Britain can thus be viewed as the western continuation of the massive movement of people that brought the CWC and steppe ancestry into central Europe a few hundred years before.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-149

Ancient genomic insights into population transitions and interactions in Neolithic Europe

Mark Lipson [1,*], Anna Szécsényi-Nagy [2], Eszter Bánffy [2], Krisztián Oross [2], Swapan Mallick [1], Nadin Rohland [1], Kristin Stewardson [1], Matthew Ferry [1], Megan Michel [1], Jonas Oppenheimer [1], Nasreen Broomandkhoshbacht [1], Eadaoin Harney [1], Susanne Nordenfelt [1], Annamária Pósa [2], Balázs Stégmár [2], Kitti Köhler [2], Balázs Gusztáv Mende [2], Tibor Marton [2], Anett Osztás [2], János Jakucs [2], Tibor Paluch [3], Piroska Csengeri [4], Judit Koós [4], Alexandra Anders [5], Katalin Sebők [5], Pál Raczky [5], Judit Regenye [6], Judit Barna [7], Mária Bondár [2], Szilvia Fábián [8], Emese Gyöngyvér Nagy [9], János Dani [9], Javier Fernández-Eraso [10], José Antonio Mujika-Alustiza [10], Carmen Alonso Fernández [11], Javier Jiménez Echevarría [11], László Márk [12], Béla Melegh [13], Joachim Burger [14], Kurt W. Alt [15], Carles Lalueza-Fox [16], Wolfgang Haak [17], David Reich [18]

[1]Department of Genetics, Harvard Medical School, Boston, United States, [2]Institute of Archaeology, Research Centre for the Humanities, Hungarian Academy of Sciences, Budapest, [3]Ferenc Móra Museum, Szeged, [4]Ottó Herman Museum, Miskolc, [5]Institute of Archaeological Sciences, Eötvös Loránd University, Budapest, [6]Dezső Laczkó Museum, Veszprém, [7]Balaton Museum, Keszthely, [8]Department of Archaeological Excavations and Artefact Processing, Hungarian National Museum, Budapest, [9]Déri Museum, Debrecen, Hungary, [10]Department of Geography, Prehistory, and Archaeology, University of the Basque Country, Vitoria, [11]CRONOS SC, Burgos, Spain, [12]Department of Biochemistry and Medical Chemistry and Szentágothai Research Center, [13]Department of Medical Genetics and Szentágothai Research Center, University of Pécs, Pécs, Hungary, [14]Institute of Anthropology, Johannes Gutenberg University, Mainz, Germany, [15]Center of Natural and Cultural History of Man, Danube Private University, Krems, Austria, [16]Institute of Evolutionary Biology, Barcelona, Spain, [17]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany, [18]Department of Genetics and Howard Hughes Medical Institute, Harvard Medical School, Boston, United States

Abstract: Ancient DNA studies have established that European Neolithic populations prior to the arrival of steppe ancestry were descended from Near Eastern migrants who admixed with resident hunter-gatherers. Many open questions remain, however, about the population dynamics of Neolithization, both in Europe specifically and as a case study of the process more generally. Using new analytical techniques and the highest-resolution genome-wide ancient DNA data set compiled to date—a total of 178 samples, 128 newly reported here, from the Neolithic and Chalcolithic of Hungary (5750–2800 BCE, n=99), Germany (5200–3200 BCE, n=42), and Spain (5400–2350 BCE, n=37)—we investigate the spatial and temporal dynamics of interactions and admixture during the Neolithic period. Admixture between farmers and hunter-gatherers occurred multiple times, beginning shortly after initial contact in each region, with predominantly local gene flow. Almost all of our sampled populations show some degree of genetic discontinuity from their predecessors, indicating that continuous gene exchange reshaped the makeup of European populations throughout the period. We synthesize our findings within a unified modeling framework but also emphasize the distinctiveness of Neolithic genetic history across the continent.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-429

PALEOGENOMES SHED LIGHT ON THE ROLE OF THE BERING LAND BRIDGE ON MIGRATIONS OF LATE PLEISTOCENE CABALLINE HORSES

Alisa Vershinina [1,*], Peter Heintzman [2], Grant Zazula [3], Mathias Stiller [4], Joshua Kapp [1], Ludovic Orlando [5], Cristina Gamba [6], Russell Corbett-Detig [1], Beth Shapiro [1]

[1]University Of California Santa Cruz, Santa Cruz, United States, [2]Tromsø University Museum, Tromsø, Norway, [3]Government of Yukon Department of Tourism and Culture, Whitehorse, Yukon Territory, Canada, [4]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, [5]Centre for GeoGenetics, University of Copenhagen, Copenhagen, [6]Biomatters ApS, Silkeborg , Denmark

Poster: Climate change drives rapid and extreme fluctuations in habitat availability and Arctic mammals are most vulnerable to these effects. Shifts in habitat availability have potentially broad evolutionary consequences by both creating and destroying barriers to dispersal for many species. Here, we explore the influence of shifting barriers to gene flow on the evolution and diversification of horses, *Equus caballus,* during the Pleistocene ice ages. We use four high-coverage, complete paleogenomes, including a new paleogenome that we isolated and assembled from a ~30,000-year-old horse metapodial found in the placer mines of Canada's Yukon Territory; the first ancient horse paleogenome from North America. Using a set of genome-wide SNPs called from this and other ancient North American and Eurasian horse genomes, we infer the demographic history of North American horses and explore the timing and nature of connectivity between horse populations on these two continents. The Bering Land Bridge allowed dispersal between Eurasia and North America during periods of cold climate and low sea level, however, a formation of mesic steppe and shrub tundra along the bridge may have obstructed migration of large grazing mammals even when the bridge was present. Our results provide new insights into the role of connectivity and barriers to gene flow in shaping genetic diversity and mitigating extinction risk.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-406

EXTENSIVE FARMING IN ESTONIA STARTED THROUGH A SEX-BIASED MIGRATION FROM THE STEPPE

Lehti Saag [1], Liivi Varul [2], Christiana Lyn Scheib [3], Jesper Stenderup [4], Morten E. Allentoft [5], Lauri Saag [1], Luca Pagani [1], Maere Reidla [1], Kristiina Tambets [1], Ene Metspalu [1], Aivar Kriiska [6], Eske Willerslev [5], Toomas Kivisild [3], Mait Metspalu [1,*]

[1]Estonian Biocentre, Tartu, [2]School of Humanities, Tallinn University, Tallinn, Estonia, [3]Department of Archaeology and Anthropology, University of Cambridge, Cambridge, United Kingdom, [4]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, [5]Centre for GeoGenetics, Natural History Museum of Denmark, Copenhagen, Denmark, [6]Department of Archaeology, Institute of History and Archaeology, University of Tartu, Tartu, Estonia

Poster: Farming-based economies appear relatively late in Northeast Europe and the extent to which they involve genetic ancestry change is still poorly understood. Here we present the analyses of low coverage whole genome sequence data from five hunter-gatherers and five farmers of Estonia dated to 4,500 to 6,300 years before present. We find evidence of significant differences between the two groups in the composition of autosomal as well as mtDNA, X and Y chromosome ancestries. We find that Estonian hunter-gatherers of Comb Ceramic Culture are closest to Eastern hunter-gatherers. The Estonian first farmers of Corded Ware Culture show high similarity in their autosomes with European hunter-gatherers, Steppe Eneolithic and Bronze Age populations, and European Late Neolithic/Bronze Age populations while their X chromosomes are in addition equally closely related to European and Anatolian/Levantine early farmers. These findings suggest that the shift to intensive cultivation and animal husbandry in Estonia was triggered by the arrival of new people with predominantly Steppe ancestry, but whose ancestors had undergone sex-specific admixture with early farmers with Anatolian ancestry.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-135

Genetic structure, diversity and variation in the populations of Flat-headed cusimanse (Crossarchus platycephalus) in Nigeria

Bukola oguntuase [1,*], Richard meisel [2] and population Genomics of ancient DNA

[1]ecotourism and wildlife management, federal university of technology, akure, Nigeria, [2]biochemistry and biology, university of houston, houston, United States

Abstract: Conservation goals have expanded from saving endangered species to sustaining biological diversity at the level of genes, individuals and populations. *Crossarchus platycephalus* occur in different populations in Nigeria, though they appear similar in their physical appearances, however, this is not enough justification for ignoring possible hidden variations in their genetic makeup. We assessed for the first time, the genetic structure and variation within and among the populations of *C. platycephalus* in Nigeria. The study was carried out on two populations of *C. platycephalus* whose habitats were separated by river Niger. Morphological assessment on individuals of the two populations showed a high level of similarity in the body measurements of the populations except in the snout and fore limbs. Loci from the sequences generated using GBS protocol was used to distinguish the populations into two clusters in the STRUCTURE analysis. In the assessment of genetic diversity, population from the west of Niger showed higher observed heterozygosity at a significant level, an indication of a richer genetic diversity within the population. Though, genetic differentiation - $F_{ST,}$ was low (0.074) in the populations, but a higher value of inbreeding coefficient - $F_{IS,}$ was recorded for populations from the east of Niger. There is a general indication that populations from the south of Niger suffer low genetic diversity and possible inbreeding depression. The outcome of this study pointed towards possibility of founder effect in the populations from the east of Niger, however further study is required in confirming this hypothesis.

Expanded summary*: Introduction

Monophyly of most cusimanse species remained unresolved even with a recent study with the use of microsatellite in the study of some cusimanse species, the monophyly of *C. platycephalus* was partially supported (sonnet *et al*., 2014), It is a gap that needs to be filled. There are no documented studies on the population structure and genetic diversity of the populations of *C. platycephalus* in Nigeria.

Study area

The study was carried out in two eco regions in Nigeria: the western part separated to the west of river Niger and the east of the river Niger also known as southern Nigeria

Morphological differentiation

The two populations were compared to each other using measurements from their body weight, total length, snout length, tail length and rump circumference, pelage was observed for variation in colour.

Molecular Analyses

DNA was extracted from the tissue of individuals from each population. Library was prepared following GBS protocol in Parchman et al., 2012

Sequencing and analysis

Sequencing was accomplished on a single lane run on an Illumina sequencer. The output of the sequencing was assembled and demultiplexed using STACKS pipeline (Catchen et al., 2013). population program in the STACKS pipeline was used to compute population-based summary statistics.

Results

Morphometric differentiation

The parameters measured were subjected to two sample t-Test in R package. All measurements were similar and not significantly different between the two populations except for the snout length and fore limb. Population from the west has shorter fore limb and longer snout

Genetic structure, diversity and variation

There were 77 loci compared for the study of genetic structure. True value of k from possible clusters was estimated according to standard methods ((Evanno *et al.,* 2005) to be two population clusters.

Keywords: Conservation, Crossarchus, Genetic diversity, Structure

Homozygosity is higher in the population from the east of Niger, while population from the west of Niger has higher observed heterozygosity. In these populations, differences observed in the mean of both heterozygosity and homozygosity are significantly different at $p < 0.0005$ (Two sample t-Test). Inbreeding coefficient is higher (0.41) in the populations from the east but not significant from the populations in the west of Niger. The estimate of F_{ST} as computed in the stacks pipeline shows little differentiation (0.074). Principle Component Analysis of haplotype diversity showed a grouping of individuals into two, Population from the east is a subset in the ellipse of the population from west. There are several shared haplotypes between the two populations.

Discussion

Morphometry of C. platycephalus

Morphological measurements are important in providing information on easily observable variations within and among populations (Linderfors *et al*., 2007). Most of the external measurements taken on *C. platycephalus* such as the mean body weight, total body length, tail length and hind limb showed no significant difference except for the snout and the fore limb which were significantly longer in one population. The mean fore limb is longer in the populations from the east which might be an indication of adaptation to the dense vegetation characterised by tropical high trees, the snout is longer in the populations from the west which could be as a result of devising means to obtaining scarce preferred diet. Adaptation as viewed by some authors (Skinner and Mitchelle, 2011; Taylor *et al*., 2011; Kingdon, 2004; Christian, 2010; Ruxton and Wilkinson, 2011) could be responsible for the variations in some of the parameters measured, more so that the populations are taken from ecological settings that are in some ways different and separated from each other.

Genetic analysis:

The structure analysis shows an evidence for two populations as confirmed by the two test for the best estimate of K according to the criteria outlined in Pritchard *et al*., (2000) and by calculating delta K (Evanno *et al.,* 2005). The structure analysis further showed evidence of migration between the two population as inferred by their ancestry in the structure analysis. A possible indication of the structure analysis presents population from the west as original population from which few individuals migrated to the east of Niger, and due to their small number and possible founder effect, are limited in genetic diversity as earlier studied in some other species in re-introduction programme (Ian, 2011).

The results of observed and expected homozygosity and heterozygosity also showed that, populations from the west have more heterozygote alleles than those from the east and the populations from the east have more homozygote alleles than those from the west. Genetic differentiation in form of F_{ST} value shows low level of genetic differentiation between the populations. Lack of genetic differentiation also confirms migration and gene flow between populations as indicated in the structure analysis. Populations from the west of Niger contained more haplotypes than population from the east of Niger, there are more unique haplotypes in western population than there are in eastern population, however a good number of the haplotypes are shared between the two populations. These findings have conservation implications, there is an indication that populations from the west are more diverse in terms of genetic make-up than those from the south. The variations observed in these genetic parameters could be as a result of reduced population size, habitat fragmentation which might restrict movement of individuals and disrupt gene flow and other factors. Even though the result on inbreeding coefficient shows no significant differences among populations, inbreeding coefficient is higher in population from the east. Generally, the outcome is a pointer to possible founder effect occurring in population from the east of Niger.

Conclusion

The morphological study and genetic differentiation of the populations of *C. platycephalus* has provided evidence of some level of differences in the population, and has also shown different levels of similarity in their physical appearance as well as their genetic component. More importantly, it has revealed the status of diversity among the populations which is an area for further investigation, there is also need for study on population genetics to further investigate founder effect in the populations, especially population from the east of Niger.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-132

The Role of Migration in Cultural Changes during the Chalcolithic period in the Levant

Eadaoin Harney [1,*], Hila May [2], Israel Hershkovitz [2], Dina Shalem [3], Nadin Rohland [4], Swapan Mallick [4], Iosif Lazaridis [4], Nick Patterson [4 5], David Reich [4 5 6]

[1]Organismic and Evolutionary Biology, Harvard University, Cambridge, United States, [2]Anatomy and Anthropology, Tel Aviv University, Tel Aviv, [3]The Institute for Galilean Archaeology, Kinneret, Israel, [4]Genetics, Harvard Medical School, Boston, [5]Broad Institute, Cambridge, [6]Howard Hughes Medical Institute, Boston, United States

Abstract: A major controversy is whether cultural change evident in the archaeological record is typically achieved through movements of people or cultural infiltration. The Chalcolithic period in the southern Levant (4^{th}-5^{th} millennium BCE) contains artifacts not detected in earlier archaeological sites of the region, yet have strong affinities to contemporary and earlier cultures from Anatolia and Iran. In order to test the hypothesis that the Chalcolithic culture of this region may have been formed through migration from the North, we analyzed new genome-wide ancient DNA data from 22 individuals from the Peqi'in cave site in Upper Galilee, Israel that are associated with the Late Chalcolithic culture of the southern Levant, thereby approximately doubling the number of samples with genome-wide ancient human DNA from the Levant. We report that that these individuals derive approximately 58% of their ancestry from populations related to those of the local Levant Neolithic, approximately 17% from populations related to the Iran Chalcolithic, and approximately 25% related to the Anatolian Neolithic, supporting the hypothesis that this population was formed in part by migration from the North. We show that population turnover continued after the Chalcolithic, as the population that the Peqi'in Cave group was a part of did not contribute to later Levantine populations from the Bronze Age, which had little or no Anatolian-related ancestry.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-136

Reconstruction of early population history of Africans in the Americas through St. Helena Island (South Atlantic) and New York City

Gretchen Johnson [1,*] on behalf of Cobb Research Lab, Fatimah Jackson [1] on behalf of Cobb Research Lab

[1]Biology, Howard University, Washington, DC, United States

Abstract: Improved sequencing technologies allow for ancient DNA research of historic remains of underrepresented groups, which can provide riveting insights into their biological history. The combination of molecular genomic analyses and morphometric anthropometric assessments, within bioarchaeological and historical contexts, give a more accurate perspective on the significant events that occurred in the past. The investigation focuses on the status of early to mid-19th century Africans on their way to enslavement in the Americas and the status of late to 18ththrough mid-19th century enslaved Africans and African Americans in New Amsterdam/New York City. This investigation reconstructs the early population history of Africans in the Americas using these two important source populations. There is an examination of the range of demographic and morphometric diversity discovered among the liberated Africans buried at St. Helena. A goal is to determine if there is any similarity with that observed among individuals found in the New York African Burial Ground. It is hypothesized that skeletal remains on St. Helena show evidence of disease and trauma similar to the skeletal remains found on the New York African Burial Ground. Furthermore, the suggestion is that the range of ancestral genomic variability found in the South Atlantic African remains is similar to the range of genomic diversity observed in the New York African Burial Ground. Together, these sites provide a continuum of insights into the early population history of these African peoples.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-134

Reconstructing relationships between historical grape varities from herbaria

Hsiao-Lei Liu [1,*], Logan Kistler [1], Oliver Smith [1], Nathan Wales [2], Jazmín Ramos Madrigal [2], Thierry Lacombe [3], Patrice This, Roberto Bacilieri [3], Thomas Gilbert [2], Beth Shapiro [4], Robin Allaby [1]

[1]School of Life Science, University of Warwick, Coventry, United Kingdom, [2]Centre for GeoGenetics, Natural History Museum of Denmark, Copenhagen, Denmark, [3]Institute National De La Recherche Agronomique, Montpellier, France, [4]Department of Ecology & Evolutionary Biology, University of California, Santa Cruz, California, United States

Abstract: Domestication mirrors the principle of natural selection. As an endemic species, the wild and the domesticated form of the European grape (*Vitis vinifera ssp. sylvestris* and *ssp. vinifera*) are scattered across the Mediterranean region. Due to overlapping habitats, recurrent introgression events between the two subspecies could have taken place before the introduction of the grafting in response to the Great French Wine Blight in 1860s caused by the pest Phylloxera (*Daktulosphaira vitifoliae*). Such introgressions could have changed the properties of commercialized grape varieties and endangered *ssp. sylvestris.*

Herbaria material offer the opportunity to track genome evolution directly through time, and so get behind introgressive events. In this poster, we aim to identify the cultivar status of each herbarium sample, the phylogenetic relationships between samples, and then to elucidate whether introgression events have re-shaped the genetic background of the wine grape through time. Historical DNA material collected from over 60 herbaria specimens covering an age range between the 16th and 18th centuries have been sequenced and annotated using shotgun sequencing, DNA target enrichmnet and GrapeReSeq database.

These data allow us to track the relationship among wine grapes and its wild relatives through time. The results will indicate how hybridization events as a whole could affect the genetic composition of a domesticated organism. It also illuminates the crucial role of museum and herbaria collections for research into recent evolutionary change.

Expanded summary*: This poster is part of the archaeogenomic project "deconstructing the origin and spread of resistance to grafting and Phylloxera". The whole work of this project focus on the interaction among wine grapes, pests, and artificial manipulations. This project is particularly interested in the genomic modifications that may have happened subsequent to the onset of grafting approaches, where Phylloxera-resistant rootstocks from North America provided a naturally selective advantage to European grape against the pest. The result will definitely the clear out the genetic component of *V. vinifera,* and contribute to amelioration of viniculture as well as the conservation strategy of the *ssp. sylvestris*.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-137

Phylogenomics and morphology of extinct birds point to Laurasian origin for paleognaths

Takahiro Yonezawa*, Takahiro Segawa, Hiroshi Mori, Jiaqi Wu, Masami Hasegawa

Abstract: The clade Palaeognathae comprises the flightless ratites and the volant tinamous, and together with the Neognathae constitutes the class Aves. It is commonly believed that Palaeognathae originated in Gondwana (the southern of two super continents of the Mesozoic era) as most of its living species are found on the southern hemisphere. However, this hypothesis has been questioned because the fossil paleognaths are mainly from the northern hemisphere in their earliest time (Late Cretaceous to Paleocene). To address the origins of Palaeognathae, we recovered nuclear genome fragments from extinct elephant birds, which enabled us to construct a reliable phylogenomic time-tree for the Palaeognathae. Based on the tree, we identified homoplasy in Palaeognathae morphological traits and reconstructed its morphology-based phylogeny including fossil species without molecular data. In contrast to prevailing theories, the fossil paleognaths from the Northern Hemisphere, such as *Lithornis*, were placed as the basal lineages. Combined with our robust divergence time estimate that enabled precise argument on the correlation with geological events, we propose a new evolutionary scenario, which contradicts the traditional idea. The ancestors of the Palaeognathae were volant birds, as estimated from their molecular evolutionary rates, and originated during the Late Cretaceous on the Laurasian continent (the northern of the two super continent of the Mesozoic era). They migrated to the Southern Hemisphere and explosively speciated around the Cretaceous-Paleogene (K-Pg) boundary. They then extended their distribution to the Gondwana-derived landmasses, such as New Zealand and Madagascar, by overseas sweepstake dispersal. Gigantism subsequently occurred independently on each landmass.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-142

The population genomics of the Neolithic and Bronze Age transitions in Western Iberia

Rui Martiniano [1 2,*], Lara Cassidy [1], Ros O Maolduin [1 3], Russell Mclaughlin [1], Nuno Silva [4], Licinio Manco [5], Daniel Fidalgo [5], Tania Pereira [5], Maria Joao Coelho [5], Miguel Serra [6], Joachim Burger [7], Rui Parreira [8], Elena Moran [8], Antonio Valera [9 10], Eduardo Porfirio [11], Rui Boaventura [8], Ana Silva [5 8 12], Chris Tyler-Smith [2], Richard Durbin [2], Daniel Bradley [1]

[1]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland, [2]Wellcome Trust Sanger Institute, Cambridge, United Kingdom, [3]Department of Archaeology, NUI Galway, Galway, Ireland, [4]Department of Genetics & Evolution , University of Geneva, Geneva, Switzerland, [5]Department of Life Sciences, University of Coimbra, [6]Palimpsesto - Estudo e Preservação do Património Cultural Lda, Coimbra, Portugal, [7]Palaeogenetics Group, Johannes Gutenberg University, Mainz, Germany, [8]UNIARQ – WAPS, University of Lisbon, Lisbon, [9]NIA - ERA Arqueologia S.A, Oeiras, [10]Interdisciplinary Center for Archaeology and Evolution of Human Behavior (ICArEHB), University of Algarve, Algarve, [11]Palimpsesto - Estudo e Preservação do Património Cultural , [12]Laboratory of Forensic Anthropology, Department of Life Sciences, University of Coimbra, Coimbra, Portugal

Abstract: We analyse new genomic data (0.05-2.95x) from 14 ancient individuals from Portugal ranging from the Middle Neolithic (4200-3500 BC) to the Middle Bronze Age (1740-1430 BC). By imputing genomewide diploid genotypes in these together with published ancient Eurasians, we can apply sensitive haplotype-based analyses to reveal new details in the integration of modern European genetic structure. While discontinuity is evident in the transition to agriculture, our results nevertheless suggest a significant subsequent local hunter-gatherer contribution to Iberian Neolithic populations. A more subtle genetic influx is also apparent in the Bronze Age, detectable from analyses including haplotype sharing with both ancient and modern genomes, D-statistics and paternal lineages. The limited nature of this introgression contrasts with the major Steppe migration turnovers within the third Millennium BCE in northern Europe, and echoes the incomplete introduction of Indo-European language to Iberia. Changes in genomic estimates of individual height across Europe are also associated with these major cultural transitions, and ancestral components continue to correlate with modern differences in stature. Lastly, we extend our analysis to a larger dataset of ancient and modern European Y-chromosome sequences to investigate the impact of past migrations on present-day male lineage diversity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-143

A genetic compendium of an island: documenting continuity and change across Irish prehistory

Lara Cassidy [1,*]

[1]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland

Abstract: Throughout the 10,000 year human occupation of Ireland, the island has witnessed several profound cultural shifts. The first four ancient Irish genomes published demonstrate that two pivotal transitions, that to agriculture and later to metallurgy, were both catalyzed by extensive population migration to the island. Moreover, both Y-chromosome and haplotype-based analyses suggest continuity between Irish Bronze Age genomes and modern Celtic-speaking populations. Adding to these conclusions, data is presented here from over 50 ancient Irish genomes (>1X) spanning from the Mesolithic to Late Iron Age (4500 BC – 500 AD). Through a combination of pseudo-diploid analyses (PCA, ADMIXTURE, D- and F- statistics) and haplotype-sharing methods, applicable through the use of genotype imputation, this dataset both confirms our published results and offers a more detailed view of the genetic processes surrounding these transitions. We explore the impact of hunter-gatherer introgression on early farming populations; the possibility of geographical and temporal structure in the Neolithic period; the complex nature by which metallurgy was introduced to the island in the Chalcolithic and Early Bronze Age; and signals of continuity between the Early Bronze and Late Iron Age periods.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

OW-POG1

Methods for combining ancient and modern genomic data sets

Montgomery Slatkin*

Abstract: Ancient DNA (aDNA) poses several difficulties for population genetic analysis. There are often low levels of endogenous DNA that can be recovered and degradation of aDNA can result in higher rates of sequencing error. In addition, for hominin aDNA, contamination from present-day humans is usually a factor. Population genetics theory can be applied to aDNA provided these features are properly accounted for and that the time dimension is incorporated into models. I will review several recent theoretical results that are relevant to the study of aDNA. (1) A theory of isolation by distance and time shows that patterns of isolation by distance are diluted and distorted by gene flow and range expansion that occurred between the times that separated two sampled genomes. (2) The projection of an ancient genome onto a present-day population provides a simple visual way to illustrate the relationship between an archaic sample and a present-day population. (3) A simple test for direct ancestry of an archaic genome to a present-day population can be developed even when aDNA coverage is so low that genotypes cannot be called.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-148

SMC2: Inferring effective population sizes and migration rates using a particle filter

Donna Henderson [1,*], Joe Zhu [1], Gerton Lunter [1]

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

Abstract: We have developed SMC^2, a method to infer the demographic history of a population by using sequential importance sampling to explicitly sample from the posterior distribution of ancestral recombination graphs (ARGs). The method is able to analyse multiple samples from several populations simultaneously, enabling the inference of both migration rates and effective population sizes as a function of time. SMC^2 is simulation-based, allowing inference under complex models, including those with asymmetric migration rates. Except for the approximation inherent in importance sampling, the method makes few additional approximations, and as a result the method has power to infer demographic parameters across a wide range of epochs.

These features make SMC^2 an attractive method for investigating ancient populations and gene flow. As such, we have incorporated ancient samples into our analyses for comparing Neanderthal and modern human population histories. In preliminary analyses of the population history of the Vindija Neanderthal, using high coverage chromosome 1 sequence data generated by Svante Pääbo's group, SMC^2 infers a long-term decrease in effective population size followed by a slight expansion around 60kya. This is consistent with the findings of Kuhlwilm *et al.* (2016), who use a Monte Carlo sample of genealogies at many loci on chromosome 21 of the Vindija Neanderthal to infer effective population sizes.

We are currently investigating historic migration by analysing high coverage samples from a selection of modern human populations, the Vindija Neanderthal, and the Denisovan. Whilst we have focused primarily on the application of SMC^2 to inferring demographic parameters, we envisage an additional utility of the posterior distribution of ARGs to identify regions of the genome of interest, such as those evolving under selective pressures.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-126

The Detection of Demographic Structure and Migration in mtDNA

Adam Benjamin Rohrlach [1 2,*], Nigel Bean [1 2], Jonathan Tuke [1], Barbara Holland [3], Wolfgang Haak [4 5], Ray Tobler [5], Alan Cooper [5]

[1]School of Mathematical Sciences, The University of Adelaide, Adelaide, [2]ARC Centre of Excellence for Mathematical and Statistical Frontiers, Melbourne, [3]School of Physical Sciences, University of Tasmania, Hobart, Australia, [4]Archeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany, [5]Australian Centre for Ancient DNA, School of Biological Sciences, The University of Adelaide, Adelaide, Australia

Abstract: The detection of demographic structure can be of key importance for assessing the modelling assumptions of a phylogenetic analysis, and may be the very question of interest when analyzing sequence data. Methods exist for the exploration of autosomal DNA, however, these methods can not be applied to non-autosomal DNA, such as mtDNA and Y-chromosomal DNA.
In this talk I will introduce a method for calculating dissimilarity distances between individuals in an alignment based on chi-square distances via Multiple Correspondence Analysis. Using these distances I will show that individuals can be tested for correlations with quantitative supplementary variables, such as geographic location, through an updated Mantel test. Finally I will describe a method for detecting possible migration routes for individuals in an alignment. To illustrate the power of these methods I will analyse the mtDNA of Aboriginal Australians obtained from hair samples with proveneance data predating European arrival in Australia.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-127

aDNA reveals a different demographic impact of the Neolithic transition in South and North of Spain

Gloria Gonzalez [1,*], Tassi Francesca [1], Kirstin Henneberger [2], Cecilio Barroso [3], Arturo Lombera [4], Ramon Fabregas [4], Aurora Grandal [5], Michael Hofreiter [2], Guido Barbujani [1]

[1]Biology, University of Ferrara, Ferrara., Italy, [2]Biology, University of Potsdam, Potsdam, Germany, [3]Archaeology, Fundacion Estudios Prehistoricos, Cordoba, [4]Archaeology, Universtiy of Santiago de Compostela, Santiago de Compostela, [5]Biology, University of A Coruna, A Coruna, Spain

Abstract: Ancient DNA (aDNA) studies are strongly contributing to shed light on a widely debated topic in archaeology, anthropology and population genetics: the dynamics of the spread of farming into Europe, or Neolithic transition. In the last few years, nuclear aDNA from prehistoric samples have led to identify the genetic legacy of the Neolithic and later human migrations into modern Europeans. Far from closing the debate, these palaeogenomes are revealing a complicated scenario, where the demographic impact of the Neolithic transition seems to have been different in different geographic areas. Describing in detail how the transition occurred in each area, and by which combination of demic and cultural changes, is now an important research priority. The Iberian Peninsula, at the western edge of major human migrations is a particular interesting area for understanding otherwise elusive aspects of European prehistory. We have taken advantage of the high percentage of endogenous DNA preserved in the petrous bone to recover nuclear genomes from Spanish prehistoric samples by shotgun sequencing. The newly generated genome data has revealed a different demographic impact of the Neolithic transition in North and South of Spain, with northern populations having a smaller Middle East component in their gene pool than southern populations. Furthermore, the palaeogenetic data has revealed prehistoric contacts between the Iberian and African populations dated back to at least 3000 years before present.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-145

Ancient DNA tracks the mainland extinction and island survival of the Tasmanian devil

Anna Brüniche-Olsen [1 2,*], Menna Jones [2], Christopher Burridge [2], Elizabeth Murchison [3], Barbara Holland [4], Jeremy Austin [5]

[1]Department of Forestry & Natural Resources, Purdue University, West Lafayette, United States, [2]School of Biological Sciences, University of Tasmania, Hobart, Australia, [3]Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom, [4]School of Physical Sciences, University of Tasmania, Hobart, [5]School of Earth & Environmental Sciences, University of Adelaide, Adelaide, Australia

Abstract: Insular species are more prone to extinction than their continental counterparts. The Tasmanian devil (*Sarcophilus harrisii*) provides a puzzling exception; surviving on the island of Tasmania for thousands of years after extinction on mainland Australia. Determining past population dynamics is fundamental for understanding why the species survived on the island of Tasmania but not on mainland Australia. Here we investigate the demographic history of the Tasmanian devil over the last ~30k years using complete and partial mtDNA genomes from 202 devils representing the extinct mainland and the extant Tasmanian populations. The mainland population had more genetic diversity than the Tasmanian population, and there was no genetic evidence for a population decline prior to its extinction, thus extrinsic and not intrinsic factors are most likely to have caused the mainland extinction.

Expanded summary*: Background

It is generally accepted that island species are more prone to extinction than their continental counterparts. Smaller population size makes island species vulnerable to demographic and stochastic processes, putting them at increased risk of negative genetic impacts—low genetic diversity and inbreeding depression [1]. In contrast, extrinsic factors such as introduced predators [2], human activities [3] or environmental change [4] are a concern for both island and continental populations, as they can affect large areas and may cause extinction before intrinsic factors impact a species [5]. There are multiple examples of island populations that have gone extinct [6], but examples of island survival and continental extinctions are comparatively rare [7].

One example of a species that went extinct across its continental range, but survived on an island is the Tasmanian devil (*Sarcophilus harrisii*). Tasmanian devils are one of the largest predators to survive the late Pleistocene (40k YBP), a period when several larger carnivores went extinct [8]. During the Pleistocene devils were widespread across mainland Australia, except the arid interior, but went extinct there during the mid-late Holocene (3.5-3k YBP) [9]. Changes in climate [10], human intensification (i.e., development of advanced tools and population size growth) [8], introduction of the dingo (*Canis lupus dingo*) [11], and multi-causal models involving two or three of these influences have been implicated for their mainland extinction.

Central to the debate around the cause(s) of the devil mainland extinction is the separation of Tasmania from the mainland by rising sea levels that formed Bass Strait around ~14k YBP [12], that isolated Tasmania from the potential impacts of dingos and human intensification. While human intensification increased on mainland Australia during the Holocene, particularly from 5–4k YBP, the human population density in Tasmania remained low and technological advances were absent [13]. The dingo, a novel top predator and competitor, was introduced to mainland Australia around 3.5k YBP [14]. Dingos spread throughout the Australian mainland, but never reached Tasmania due to the Bass Strait barrier. Thus, the presence of human intensification and dingos on the mainland, but absence of these factors in Tasmania, has provided strong circumstantial evidence that one or both were the major contributor to the mainland extinction of devils and thylacines.

The devil population in Tasmania went through extensive bottlenecks during a period of changing climate following increased ENSO activity 5–3k YBP [15]. This bottleneck likely resulted in the observed low genetic diversity in the extant population at mitochondrial [16] and nuclear loci [16]. Knowledge of the demographic history of the mainland devil population is limited. Ancient DNA (aDNA) analysis of immune genes has shown that haplotypes from the extinct mainland population were embedded within the current Tasmanian diversity [17], suggesting a small long-term population size or a species-wide late Pleistocene or early Holocene population bottleneck.

Objectives
Here we trace the demographic history of the Tasmanian devil and investigate the dynamics of mainland and island devil populations to explore whether levels of genetic diversity were similar, and if the mainland experienced a gradual rather than an abrupt decline prior to its extinction. Specifically we address the following questions:

- Was genetic diversity of the extinct mainland population different to the levels found in the extant devil population?

- When did the devils on the mainland and Tasmania diverge?

- Did the two devil populations have similar effective population size trajectories prior to the mainland devils extinction?

Methods
Using mitochondrial genome sequences from 202 Tasmanian devils from both the extinct and the extant population we conduct a series of genetic analyses to infer the species demographic history. The samples are from sub-fossil bones, historical museum specimens, and modern tissue samples—these date from the present to 17k YBP. Briefly, we estimated genetic diversity [18], tested for temporal signal in order to estimate a mitochondrial rate estimate [19], and inferred the populations demographic histories using BEAST [20] and Approximate Bayesian Computation [21].

Significance
This study traces the unusual case of mainland extinction and island survival of a species. Our results shed light on the demographic history of the world's largest extant marsupial carnivore and contribute new knowledge to the mainland devil extinction debate. By combining mtDNA sequences from the extinct mainland and the extant Tasmanian populations we show that the mainland populations were much larger than the island population, with only the Tasmanian population surviving to the present day, albeit with reduced effective population size and diversity.
Contemporaneous extinction on the mainland and decline in Tasmania combined with the absence of human intensification and dingos in Tasmania, suggest a common factor to their population declines. Given the pronounced decline in Tasmania, combined with the greater severity of ENSO climate change on the mainland, we consider these climate changes, probably in synergy with human intensification and/or the establishment of the dingo population, to have contributed to the extinction of devils on the mainland.

References

1. Frankham R: **Inbreeding and extinction: Island populations**. *Conserv Biol* 1998, **12**(3):665-675.
2. Sax DF, Gaines SD: **Species invasions and extinction: The future of native biodiversity on islands**. *Proceedings of the National Academy of Sciences* 2008, **105**(Supplement 1):11490-11497.
3. Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM: **Accelerated modern human–induced species losses: Entering the sixth mass extinction**. *Science Advances* 2015, **1**(5).
4. Cooper A, Turney C, Hughen KA, Brook BW, McDonald HG, Bradshaw CJ: **Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover**. *Science* 2015, **349**(6248):602-606.
5. Spielman D, Brook BW, Frankham R: **Most species are not driven to extinction before genetic factors impact them**. *P Natl Acad Sci USA* 2004, **101**(42):15261-15264.
6. Burbidge AA, Manly BF: **Mammal extinctions on Australian islands: causes and conservation implications**. *J Biogeogr* 2002, **29**(4):465-473.
7. Guiler ER: **Thylacine: the tragedy of the Tasmanian tiger**: Oxford University Press; 1985.
8. Johnson C: **Australia's mammal extinctions: a 50000 year history**: Cambridge University Press; 2006.
9. Gould RA, O'Connor S, Veth P: **Bones of contention: reply to Walshe**. *Archaeol Ocean* 2002:96-101.
10. Brown OJF: **Tasmanian devil (*Sarcophilus harrisii*) extinction on the Australian mainland in the mid-Holocene: multicausality and ENSO intensification**. *Alcheringa* 2006, **30**(sup1):49-57.
11. Letnic M, Greenville A, Denny E, Dickman CR, Tischler M, Gordon C, Koch F: **Does a top predator suppress the abundance of an invasive mesopredator at a continental scale?** *Global Ecol Biogeogr* 2011, **20**(2):343-353.
12. Lambeck K, Chappell J: **Sea level change through the last glacial cycle**. *Science* 2001, **292**(5517):679-686.
13. Lourandos H: **Continent of hunter-gatherers: new perspectives in Australian prehistory**: Cambridge University Press; 1997.
14. Gollan K: **The Australian dingo: in the shadow of man**. *Vertebrate zoogeography and evolution in Australasia* 1984:921-927.
15. Brüniche-Olsen A, Jones ME, Austin JJ, Burridge CP, Holland BR: **Extensive population decline in the Tasmanian devil predates European settlement and devil facial tumour disease**. *Biol Letters* 2014, **10**(11).

16. Miller W, Hayes VM, Ratan A, Petersen DC, Wittekindt NE, Miller J, Walenz B, Knight J, Qi J, Zhao FQ *et al*: **Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil)**. *P Natl Acad Sci USA* 2011, **108**(30):12348-12353.
17. Morris K, Austin JJ, Belov K: **Low major histocompatibility complex diversity in the Tasmanian devil predates European settlement and may explain susceptibility to disease epidemics**. *Biol Letters* 2013, **9**(1).
18. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data**. *Bioinformatics* 2009, **25**(11):1451-1452.
19. Duchêne S, Geoghegan JL, Holmes EC, Ho SYW: **Estimating evolutionary rates using time-structured data: a general comparison of phylogenetic methods**. *Bioinformatics* 2016:btw421.
20. Drummond AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees**. *Bmc Evol Biol* 2007, **7**.
21. Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin J-M, Estoup A: **DIYABC v2. 0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data**. *Bioinformatics* 2014:btt763.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-146

Model-based inference of gene flow using present-day and ancient hominin genome samples

Diego Ortega-Del Vecchyo [1,*]

[1]Department of Integrative Biology, University of California, Berkeley, United States

Abstract: The inference of past demographic events using genomic information of human fossils along with present-day DNA information is challenging due to the particular features of ancient DNA. Four important characteristics when performing inferences using ancient DNA are the age difference between different fossils and present-day samples, the possibility of low genomic coverage, the presence of contamination, and sequencing errors. Here I introduce a likelihood-based method that takes those characteristics into account to estimate population sizes, divergence times and gene flow between pairs of populations in an isolation-with-migration model. I mostly focus on how the method estimates gene flow parameters. The inference method uses the number of sites that have a particular genotype configuration in two sampled individuals, one from the present-day and one from a fossil. I model the changes in the allelic state that arise due to contamination and sequencing errors by using information from sequencing reads. The age of the fossil, and the uncertainty of this parameter, is also modeled. To find the probability of each genotype configuration given the demographic parameters I calculate the product of the probability of each genotype configuration given each topology, and the probability of each topology over the space of all possible topologies. I test the performance of this method using simulations and explore how the misspecification of the demographic model affects the results.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

OW-POG3

A palaeogenomic perspective of near-extinction population dynamics

David Díez-del-Molino [1,*], Patricia Pečnerová [1], Love Dalén [1]

[1]Deptartment of Bioinformatics and Genetics, Swedish Museum Of Natural History, Stockholm, Sweden

Abstract: Processes leading up to species extinction are typically characterized by habitat fragmentation and decline in geographic distribution, both related to a reduction in population size. With the transition from traditional ancient DNA methods to palaeogenomics we have now the tools to assess and characterize the genomic responses to dramatic demographic declines and to identify signatures of genetic erosion across such declines by using serial sampled data to track genome-wide changes in real time. We present evidence that the expected loss of genome-wide heterozygosity and increase in inbreeding due to small population size in genomes sampled near-extinction are accompanied by an accumulation of genomic deletions and detrimental mutations, thus following the expectation of relaxed purifying selection. These results provide incomparable testimony of the value of datasets integrating heterochronous complete genomes at allowing to characterize evolutionary processes, such as near-extinction genomic dynamics, in real time and can be useful to better understand modern-day species facing similar threats as extinct species did during their final demise.

Expanded summary*: In a context of increased extinction rates, a growing interest in the role of population demographics in species disappearance is little surprising. The widespread concern that a small population size will lead to higher extinction rates in endangered species stems from the *small population paradigm* which postulates that a small population size in itself can lead to an increased risk of extinction due to stochastic and genetic processes. However, whether these genetic factors have a direct impact on pre-extinction dynamics is unclear and thus a detailed estimation of temporal changes in genome-wide diversity as well as reconstruction of the species' demographic history leading up to its extinction are required. At the same time the genetic burden of detrimental mutations in endangered species have generally received little attention, and thus the rate and proportion of accumulation of deleterious variants and their role in the disappearance of species on the brink of extinction are largely unknown. Also, interpreting the levels of inbreeding and detrimental mutations in modern-day threatened species is commonly restricted to comparisons among different species due to the absence of temporal baseline estimations.

In this project we offer a palaeogenomic perspective of these questions, and thus we circumvent the limitations of modern datasets by integrating heterochronous genomic information from samples of known age. This allows us to study evolutionary processes by tracking genome-wide changes, such as decreased heterozygosity, increased inbreeding, and accumulation of genomic deletions and detrimental mutations in populations subjected to characteristic demographic declines, in real time. Our results provide incomparable testimony of the value of datasets integrating heterochronous complete genomes at allowing to characterize and test hypotheses about the evolutionary processes that shape near-extinction genomic dynamics, and can be useful to better understand modern-day species facing similar threats as extinct species did during their final demise.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

OW-POG5

Paleo-population genetics: Illuminating the role of selection in shaping human diversity

Yassine Souilmi [1,*], Raymond Tobler [1], Matthew Robinson [2], Wolfgang Haak [3], Matthew Williams [1], Christian Huber [4], Fernando Racimo [5 6], Lars Fahren-Schmidt [7], Kirk Lohmueller [4], Iain Mathieson [8], Peter Visscher [2], Alan Cooper [1]

[1]Australian Centre for Ancient DNA, University of Adelaide, Adelaide, [2]Queensland Brain Institute, University of Queensland, Brisbane, Australia, [3]Department for Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany, [4]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, [5]Department of Biological Sciences, Columbia University, [6]The New York Genome Center, New York, [7]Anthropology Department, University of California, Santa Cruz, Santa Cruz, [8]Department of Genetics, Harvard Medical School, Boston, United States

Abstract: Ancient human DNA studies have primarily focused on the questions surrounding ancestral migrations, admixture, phylogeography, and demography. To date, however, these studies have not fully explored the role of adaptation in shaping modern human diversity. With the number of publicly available ancient human genomes likely to hit the 1000 mark in 2017, it is now possible to utilise a range of population genetic approaches to elucidate the genetic history of multiple human phenotypes.

Using a comprehensive aDNA database, the Online Ancient Genome Repository (OAGR; oagr.org.au), we utilised ~300 ancient Eurasian genomes to explore human phenotypic evolution from the late Pleistocene to the present, a period covering the major socio-cultural transitions in human history. For each sample, we predicted 40+ putatively fitness-associated traits, including many common diseases, using a novel method that accounts for population structure to produce unbiased estimates of SNP effects. By assessing differences among groups and along spatiotemporal clines against a null model of drift, we were able to infer the specific times and places that selection has shaped modern Eurasian phenotypic diversity. Further, by performing joint trait analyses, our method allowed us to examine the extent to which modern Eurasian phenotypes have co-evolved; i.e., whether correlated selection has simultaneously affected multiple phenotypes.

This analysis represents the first step in building a comprehensive spatiotemporal map of human adaptation over periods covering the major socio-cultural transitions in human history, providing a window into the factors that have shaped modern human diversity and pathology.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-141

Exploring Archaic Introgression in Modern Human Populations

Evelyn Jagoda [1,*], Daniel Lawson [2], Terence Capellini [1], Luca Pagani [3]

[1]Human Evolutionary Biology, Harvard Univserity, Cambridge, United States, [2]Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom, [3]Estonian Biocentre, Tartu, Estonia

Abstract:

Recent studies have reported evidence suggesting that portions of contemporary human genomes that were introgressed by archaic hominin populations went to high frequencies due to positive selection. However, no study to date has specifically addressed the post-introgression population dynamics of these putative cases of adaptive introgression. Here for the first time, we define and distinguish between two different cases of adaptive introgression. The first is the classic notion of adaptive introgression (AI), in which archaic haplotypes rose to high frequencies in humans as result of a selective sweep that occurred shortly after the adaptive introgression event. This is distinct from a second definition of selection on standing introgressed variation (SI) in which an introgressed haplotype initially segregated neutrally and subsequently underwent positive selection. Using a geographically diverse dataset, we report novel cases of both types of selection on introgressed variation with potential biological relevance. However, we find a lack of enrichment of AI signal across our data and find few detectable cases of AI that are widespread throughout Eurasia. We are now actively pursuing these and other cases of AI and SI using high-throughput computational and experimental wet lab assays, with the goal of uncovering the variants driving each selection signal, and understanding the selected phenotypes that drove these haplotypes to high frequency.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Population genomics of ancient DNA

POB-147

Inferring population split times and rates of gene flow between populations from multiple genome sequences

Ke Wang [1,*], Stephan Schiffels [1]

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

Abstract: Given the rising number of complete human genome sequences worldwide, many novel methods have been developed for studying population history. One of the most widely used tools for analyzing the ancestral relationship between populations from genome sequence data is the Multiple Sequentially Markovian Coalescent (MSMC). MSMC fits a model of piecewise constant population sizes and cross coalescence rates to the observed pattern of mutations in multiple phased genomes. While MSMC has the advantage of being very flexible while making only few a priori assumptions, it cannot explicitly estimate split times and gene flow between populations. Here, we present an extension to MSMC, which estimates these parameters explicitly. Specifically, we fit an Isolation-Migration model to the estimated distribution of coalescence times within and across two populations, computed from MSMC's piecewise constant model of coalescence rates. Our method allows explicit estimation of demographic parameters like migration rates, population sizes and the population split time directly from MSMC's output. Our approach provides a way to bridge the gap between the full flexibility and high complexity in piecewise constant models such as in MSMC, and the low complexity and easier interpretability of parameterized models, as implemented in many other inference tools. The method provides a useful extension to MSMC and will be relevant for many population genomic studies.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

OT-PM2

Mutation and horizontal gene transfer as drivers of short-term adaptation within the gut microbiota

Isabel Gordo*, Nelson Frazão [1], Ana Sousa [2], Michael Lassig [3]

[1]Evolutionary Biology Group, Instituto Gulbenkian de Ciência, Oeiras, [2]ibimed, University Aveiro, Aveiro, Portugal,

[3]Physics Department, University of Cologne, Cologne, Germany

Abstract: The mammalian intestine is home for many bacterial species that collectively shape the physiology and health of their hosts. Our knowledge about how each species inhabiting the gut evolves is still very limited. We have been using commensal Escherichia coli strain colonizing mice to study how mutation and horizontal gene transfer contribute to the emergence of strain diversity within the microbiota. Through experimental evolution, next generation sequencing, phenotypic assays and microscopy, we reveal the rapid occurrence of multiple HGT events, which are mediated by phages from the resident microbiota to the colonizing E. coli. We find that the ecological scenario found by the colonizing lineage is a key determinant of the mechanism of its future evolutionary change. In particular, E. coli evolution is dominated by phage-mediated HGT when its direct competitors are present in the microbiota at high abundance, otherwise it is dominated by mutation accumulation. Furthermore we demonstrate that the genome integration of incoming phages by the new E. coli colonizer results in immunity against phages produced by its resident competitors, allowing coexistence between phylogenetic groups A and B1 strains of E. coli in the mouse gut.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-311

The order of the adaptive steps of Escherichia coli in the mammalian gut is influenced by a mutational hotspot and strong effect mutations

Ana Sousa*, Marta Lourenço [1], Ricardo S. Ramiro [2], Daniela Güleresi [2], Joao Barroso-Batista [2], Karina B. Xavier [2], Isabel Gordo [2]

[1]Pasteur Institute, Paris, France, [2]Gulbenkian Institute of Science, Oeiras, Portugal

Abstract: Strong selective pressures on model organisms are typically used to study evolution in real time. The impressive level of parallelism in the adaptive process is a possible consequence of this design and therefore its relevance to natural ecosystems is still debatable.

Here we uncover the repeatability of evolution during the colonization of the mammalian gut by a commensal bacterium as it adapts to its natural environment, where multiple selective pressures exist. As an adaptive walk is expected to involve more than one step, we have studied the adaptive mutations corresponding to the first and second steps of adaptation of *Escherichia coli* to the mouse gut. We observed that the order of the adaptive steps was strongly affected by a mutational hotspot for the first step with an exceptionally high mutation rate of 10^{-5}. Following this first step, which consisted of inactivating a metabolic operon, one third of the subsequent adaptive mutations were found to have a selective effect as high as the first.

By discretizing the adaptive steps we aimed to understand not only the rhythm but also the repeatability of adaptation in the gut environment. Whole genome sequencing of clones isolated from different mice revealed 7 parallel targets comprising: three membrane transporters, one repressor of a metabolic operon, one major regulator involved in the aerobiosis/anaerobiosis transition, one large duplication, and a protein of unknown function. Interestingly all mutations tested, which occurred in regulatory regions, either up-regulated the targeted gene directly, or downstream genes in the regulatory cascade.

One possible explanation for the order of events could result from the second mutations being weaker or even deleterious in the ancestral genotype. In fact we estimated that at least one third of the second step mutations were as strong as those responsible for the first step, showing that strong effect mutations were still available for adaptation. However, many showed a smaller average effect. Therefore, it is possible that the large effect size of the first-step may have contributed for the order observed. A much more likely explanation can, however, be provided by differences in mutation rate.

The pattern of polymorphism emerging in the evolving bacterial populations was characterized by periodic selection, which reduced diversity, but also frequency-dependent selection, actively maintaining genetic diversity. Furthermore, the continuous emergence of similar phenotypes due to distinct mutations, known as clonal interference, was pervasive.

Evolutionary change within the gut is therefore highly repeatable within and across hosts, with adaptive mutations of selection coefficients as strong as 12% accumulating without strong constraints on genetic background. In vivo competitive assays showed that one of the second steps exhibited positive epistasis with the first, while another exhibited negative epistasis. The data shows that strong effect adaptive mutations continuously recur in gut commensal bacterial species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-312

A journey into the blood microbiome of Biomphalaria snails, intermediate hosts of the human blood-fluke S. mansoni

Winka Le Clec'h [1,*], Tim J. C. Anderson, Frédéric D. Chevalier [1]

[1]Genetics, Texas Biomedical research Institute, San Antonio, United States

Abstract: The microbiome – the microorganism community that inhabits on or within animal body – is increasingly recognized shape many aspects of its host biology and is a key determinant of health and disease. The central aim of this study is to investigate the microbiome of aquatic snails (genus *Biomphalaria* sp.) that are the intermediate host for the human blood fluke parasite (*Schistosoma* sp.), causative agent of a parasitic disease that infects ~67 million people in sub-Saharan Africa and South America. In this work, we characterized the diversity and abundance of microorganisms within the *Biomphalaria* hemolymph (*i.e.* blood) using next generation sequencing (Illumina MiSeq) of the V4 variable region of the bacterial 16S ribosomal DNA. We characterized the microbiome composition of hemolymph from 5 snails representing 7 different populations of *B. glabrata*, and one population of *B. alexandrina*. We observed (i) 554-6,976 operational taxonomic units (OTUs) per snail (total 76,097), where OTUs are defined as sequences differing by >3% and (ii) significant differences in microbiome composition at the level of individual snails, snail population and snail species. Moreover, we hypothesize that the microbiome may represent a critical, but unexplored intermediary in the snail-schistosome interaction. Indeed hemolymph is in very close contact with the parasite at each step of its development. To investigate this aspect of the host-parasite interactions, we will characterize the microbiome of *S. mansoni* infected snails across the parasite life-cycle.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

OT-PM5

Dynamics of within- and between-host evolution in the gut microbiome

Benjamin Good [1,*], Nandita Garud [2], Katherine Pollard [2 3], Oskar Hallatschek [4]

[1]Departments of Physics and Bioengineering, University of California Berkeley, Berkeley, [2]Gladstone Institutes, [3]Institute for Human Genetics, Institute for Computational Health Sciences, and Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, [4]Departments of Physics and Integrative Biology, University of California Berkeley, Berkeley, United States

Abstract: The composition of the gut microbiome can vary dramatically based on host species, diet, and the identities of other co-colonizing strains. Yet the evolutionary forces responsible for maintaining this high degree of specialization remain poorly understood. With the falling costs of whole-genome metagenomic sequencing, it is now possible to resolve genetic variation within individual gut bacterial species, which provides new opportunities to learn about the population genetic processes that shape the microbiome. Here, we analyze patterns of single nucleotide polymorphisms across a large panel of metagenomic samples from the Human Microbiome Project. These data provide information about short-term evolution within hosts, as well as longer-term evolution across the host population. In many species, we observe an overall separation of timescales between within- and across-host diversity, which still varies dramatically among individual hosts. Some individuals are colonized by a few strains as diverse as the panel-wide pool, so that within-host diversity is dominated by differences in the frequencies of the strains. Other hosts are more consistent with a single dominant strain, so that bacterial genotypes can be approximated by an extra “haploid” chromosome for each host/species combination. This allows us to estimate site frequency spectra to study natural selection, as well as to phase metagenomes to study patterns of intragenic recombination within individual bacterial species. In a subset of time-resolved samples, this approximation also provides a more finely resolved view into how these genomes change over the lifespan of a single host.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-309

Wolbachia abundance in Drosophila melanogaster decreases as host ages

Zhenglong Gu [1], Yuan Si [1,*]

[1]Cornell University, Ithaca, United States

Abstract: Wolbachia is a common endosymbiont widely distributed in arthropod species, including Drosophila melanogaster. It is best known for the ability to increase its prevalence by altering host reproductive systems. Recent findings revealed that wolbachia can affect the lifespan of drosophila in a manner depending on the genotypes of both wolbachia and the host. However, it is unclear how the abundance of wolbachia fluctuates along the process of aging in host. To investigate the roles of wolbachia in the host aging, we examine the pattern of wolbachia abundance in Drosophila melanogaster over time. We measure both relative and absolute abundance of wolbachia. The relative abundance was measured using 16s rRNA sequencing of gut microbiome, while the absolute abundance by Droplet Digital PCR (ddPCR). The guts of flies were dissected for DNA extraction at four different time points: 10, 20, 30 and 40 days. Our preliminary results from 16s rRNA sequencing showed that the relative abundance of wolbachia decreased along the process of aging and it is positively associated with species diversity in the gut microbiome. More research is undergoing to investigate the interaction between wolbachia and gut microbiome, to quantify the absolute abundance of wolbachia over time, and eventually to elucidate the role of wolbachia in Drosophila aging.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-323

Is there a link between aging and microbiome diversity in exceptional mammalian longevity?

Graham Hughes [1,*], John Leech [2], Sébastien Puechmaille [3], Jose Lopez [4], Emma Teeling [1]

[1]Biology and Environmental Science, University College Dublin, Dublin, [2]School of Microbiology, University College Cork, Cork, Ireland, [3]Zoology Institute, Greifswald University, Greifswald, Germany, [4]Halmos College of Natural Sciences and Oceanography, Nova Southeastern University, Dania Beach, Florida, United States

Abstract: A shift in microbiome composition has been linked to biological aging in mice and humans, suggesting a role of gut flora in pathogenic aging phenotypes, negatively affecting overall life and health-span. Whether age-related microbial shifts are observed in non-model organisms or are a symptom rather than a causative mechanism of ageing has yet to be conclusively answered. Bats are exceptional mammals not only because of their capability of powered flight but also due to their diverse range of life histories, with exceptional longevity demonstrated in a number of species. One such longevity specialist, *M. myotis*, can live up to 37 years given its body mass (28.55g), compared to the maximum 4-year lifespan in mice of similar size (20.5g). For the first time, using the 16s rRNA gene, we have investigated the anal microbiome of *Myotis myotis* with respect to aging, using a population of over 50 individuals. When comparing juvenile to adult bats, we find no shift in bacterial phyla abundances, as has been observed in other mammals. Metagenomic analysis highlights pathways involved in metabolism, energy consumption, DNA repair and oxidative phosphorylation, all of which can play an important role in both aging and flight. We also highlight the high abundances of 'Proteobacteria' in *M. myotis*, as observed in other bats, implying a 'bat-specific' microbiome relative to non-chiropteran mammals. Our analyses highlight the need for, and utility of, using long lived non-model organisms to elucidate and understand the microbiome-aging link further.

Expanded summary*: The importance of the gut microbiome in facilitating the fermentation of nutrients, such as carbohydrates, into short chain fatty acids for use by the host is well known. In addition to a role in nutrition, more evidence is accumulating pointing to "microbiome-wide associations" to health and disease in humans and other hosts, linking gut flora to aging. Biological aging is characterized by the progressive decline of function, increased frailty and an increase in chronic disease. Studies of the human microbiome have reported shifts in microbial composition across different stages of life, with a high degree of variability at the two extremes of infancy and old age compared to relative stability during adulthood. Similar microbial shifts are observed in mice, such as the decrease in bacteria that synthesize vitamin B12 in older age cohorts, leading to overall changes in microbiome composition and function in frailty. However the question of whether microbial shifts are a symptom rather than a driver of aging has yet to be conclusively answered. Bats are exceptional mammals not only because of their capability of powered flight and high metabolism but also due to the diverse range of life histories they exhibit, with exceptional longevity being of particular interest. Within bats, a number of species in the family Vespertilionidae have demonstrated extreme longevity given their body size and metabolic demands. Surviving in the wild requires maintaining agility, speed and high frequency hearing, to both capture prey and evade predation on a daily basis. Therefore, a long lifespan in bats coincides with a long health-span. Elucidating the changes that occur in microbial composition over time in these exceptionally long-lived species will shed light on the role of the microbiome in extended health-spans. In this study, we have used the 16S rRNA gene and high-throughput sequencing to characterize the structure and function of the anal microbiome in more than 50 wild, exceptionally long–lived insectivorous *Myotis myotis* bats (maximum lifespan 37 years) using a non-lethal sampling method. This large sample size allows robust analyses of the microbiome for this species. Ages ranged from 0–4+ years old, containing both juvenile and adult bats. Bacterial phyla across all samples were characterized and quantified using the QIIME software package and the Greengenes database, while putative genes expressed by bacteria in the bat anal microbiome were inferred using the PICRUSt software. Our results highlighted an abundance of Proteobacteria in *M. myotis*, showing similarities to other bat microbiome studies, and a 'bat-specific' signature compared to other mammals. Principal Co-ordinates Analysis, Analysis of similarities and Kruskal-Wallis tests found no significant differences or microbial shifts between juvenile and adult bats when comparing microbiome composition, contrasting to other mammalian microbiome studies, highlighting a potentially static bat microbial community. KEGG pathway analysis of gene expression in bacteria found in the *M. myotis* microbiome using PICRUSt indicates an abundance of genes involved in metabolism, energy processing, DNA repair and oxidative phospohrylation, all of which may play a role in sustained, powered flight. Given the high rate of metabolism observed in bats, a microbiome contributing to

metabolism and oxidative damage would provide both a physiological and adaptive advantage. This analysis of the microbiome of the exceptionally long-lived bat, *M. myotis*, suggests a potentially unchanging microbiome, contributing to the demands of a high-metabolic lifestyle, which does not undergo the age-linked shifts observed in other mammals. This work utilizes a non-model organism, using non-lethal sampling methods, allowing for longitudinal studies and future research. This research incorporates aging data, microbial abundance data and a comparative biology approach to investigating the 'aging-microbiome' association, and is of broad interest to the membership of SMBE, and a wider fields of microbiology, comparative biology and gerontology.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-308

Comparative analysis of free range cattle and human semen microbiomes.

Crystal Hepp*, Jill Cocking [1]

[1]School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, United States

Abstract: In this abstract, we present a preliminary microbiome analysis of seminal fluid collected from free range beef cattle. Importantly, seminal fluid is essential for reproduction, and there is a wide range of bacterial species that have been associated with human male infertility. While beef cattle fertility is not a matter of public health, necessarily, the industry has an $88.25 billion impact on the United States economy. A veterinarian donated five Red Angus bull semen samples from individuals belonging to the same 160 head free-range beef cattle herd, and we extracted DNA and performed microbial community amplicon sequencing (region V4 of 16S rRNA) according to the Earth Microbiome Project protocol. This was done as part of a more holistic herd microbiome study, where corral soil, fecal, milk, and vaginal samples were also included. A principal co-ordinates analysis revealed that the bull semen microbiome is, not surprisingly, distinct from the other sampled communities. Bacteria within the family *Pasteurellaceae* were the most abundant across the five individuals. Bacterial genera that were present among all five samples included *Fusobacterium*, *Campylobacter*, *Oscillospira*, and *Leucobacter*. Interestingly, in human males, *Fusobacterium periodonticum* is associated with higher semen volume. Finally, we found *Ureaplasma* in four of five individuals, *Mycoplasma* in three individuals, and *Parachlamydiaceae* in one individual. All three of these organisms are known to cause fertility issues in either humans or other animal species, and we are in the process of collecting health data regarding the five bulls. We will compare our survey to publicly available human seminal fluid 16S rRNA sequences to better understand niche preference and bacterial strain differences in this environment.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-322

Single cell genomics of malaria infections

Simon Trevino [1,*], Standwell Nkhoma [1], Shalini Nair [1], Benjamin Daniel [2], Karla Moncada [2], François Nosten [3], Ian Cheeseman [1]

[1]Texas Biomedical Research Institute, [2]University of Texas Health Science Center at San Antonio, San Antonio, United States, [3]Shoklo Malaria Research Unit, Mae Sot, Thailand

Abstract: Single-cell genomics can provide a means to determine the genetic structure of complex communities of unicellular organisms. Genetic analysis of infections with malaria parasites are complicated by multiple parasite lineages. These cannot be unambiguously determined from bulk resequencing efforts, severely restricting the effectiveness of association studies, and of our understanding of gene flow through parasite populations. To better understand parasite ecology at the level of individual cells, we have developed a method to capture single-cell haplotypes. Our optimized approach uses fluorescence assisted flow cytometry to capture singly-infected red blood cells, followed by whole genome amplification and next generation sequencing. By focusing our single cell method on replicating parasites we improve the overall success rate of amplification reactions (from 65% to 95%) and routinely generate near complete capture of the 23Mb parasite genome (mean genome coverage 85%). We use this approach to sequence hundreds of single genomes from an area of intense malaria transmission in Chikhwawa, Malawi. This data allows us to build a fine scale picture of the population structure of malaria parasites, both within and between infections.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-314

Gut microbiome analysis of diverse sub-Saharan African populations reveals a spectrum of Prevotella dominated bacterial compositions

Matthew Hansen*, Meagan Rubel [1], Aubrey Bailey [2], Jaanki Dave [3], Alessia Ranciaro [1], William Beggs [1], Kyle Bittinger [4], Frederic Bushman [2], Sarah Tishkoff [1]

[1]Genetics, [2]Microbiology, [3]University of Pennsylvania, [4]Pediatrics, Children's Hospital of Pennsylvania, Philadelphia, United States

Abstract: The human gut microbiome (GM) is an interdependent community of microbes that has a large role in the nutrition, metabolism, and immune response of the human host. "Western" diets typical of industrialized populations correlate with high abundance of *Bacteroides*, while the GM of hunting and gathering societies correlate with high abundance of *Prevotella*. Here we report the GM bacterial composition, based on bacteria ribosomal sequence (16S rRNA V1-V4) classification, from 7 populations in Tanzania and Botswana, which include traditional pastoralists, hunter-gatherers, and small-scale rural agriculturalists. We also include individuals from Philadelphia as a Western comparison group. We observe a *Bacteroides/Prevotella* gradient, with the Hadza at the high *Prevotella* end of the spectrum and the Western samples at the high *Bacteroides* end. Notably, we find that a subset of rural Bantu agriculturalists of Botswana share similar GM composition to the Western samples, including the predominance of the same *Bacteroides* operational taxonomic unit (OTU). We also find compositional differences between sexes for several populations, including the Hadza. In general we observe that although the African cohort is marked by an abundance of *Prevotella*, there is a range of GM compositions that vary significantly among populations that is not solely captured by differences in subsistence practice. Analysis of genome-wide SNP array data that genetic similarity is more strongly associated with abundance-weighted GM composition (weighted unifrac) than geographic similarity, and this association is weaker when GM compositional distance is measured using bacterial presence/absence (unweighted unifrac).

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-415

PATTERNS OF PURIFYING SELECTION IN THE HUMAN MICROBIOME

Nandita Garud [1,*], Benjamin Good [2], Oskar Hallatschek [2], Katherine Pollard [1]

[1]Gladstone Institutes, Ucsf, San Francisco, [2]University of California Berkeley, Berkeley, United States

Poster: The human gut microbiome plays an important role in host physiology. Understanding the selective forces shaping nucleotide-level diversity in the microbiome can reveal how bacteria are adapted to the host and allow us to predict how a bacterial population might respond to rapid changes in host diet and health.

We examined SNP-level diversity in metagenomic data from a panel of 180 healthy North American individuals from the Human Microbiome Project. To make inferences about selection, we first controlled for population substructure for each bacterial species within and between hosts. For most species, SNP-level diversity within several hosts is low, suggesting that these hosts harbor a single dominant strain. We call these 'haploid hosts'. Based on patterns of fixed differences at synonymous and non-synonymous sites between the haploid hosts for many species, we estimate that purifying selection is weak and has played an important role in shaping nucleotide diversity. For example, we observe an excess of rare non-synonymous variants relative to rare synonymous variants from haploid hosts and depressed linkage disequilibrium between non-synonymous sites relative to synonymous sites. Additionally, within-host time samples less than 1 year apart show few SNP changes, suggesting that purifying selection plays a role in maintaining the constancy of the microbiome despite hosts experiencing daily rapid fluctuations in diet and metabolism.

The extent of purifying selection and lack of evidence for adaptation derived from SNP changes observed in natural populations of bacteria in the human microbiome is unexpected compared to the rapid rate of adaptation observed in experimental evolution of bacteria. The notion that large census population sizes should be conducive to seeing many beneficial mutations does not seem to apply to the most prevalent bacteria we harbor in out guts. Understanding why bacteria in our guts are so resistant to change will help us link the role that bacteria play in our health.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

OT-PM1

The Extraordinary Evolution of the Human Microbiome

Howard Ochman [1,*]

[1]Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: Despite the large body of work concerning the human microbiome and its role in human health, there is relatively little information about how the microbiome evolves or the factors causing differentiation among species. Early analysis of the gut microbiomes of great ape species, including chimpanzees, gorillas and humans, revealed that the phylogeny based on microbiome compositions is congruent with the known relationships of the hosts. Our investigations of the microbiomes of great apes have informed several features of the human microbiome, such as the existence of enterotypes (*i.e.*, microbial communities having discrete species compositions), and the effects of diet, provenance, disease state and social structure on microbiome composition. By comparing the gut microbiomes of great ape species in a phylogenetic context, we have shown that human gut microbiomes have loss substantial amounts of bacterial and archaeal diversity over the course of great ape diversification. Moreover, by tracing the evolution of specific bacterial strains within great apes, we show that multiple lineages of the predominant bacterial taxa in the gut have co-diversified with their hosts over the past 15 million years.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-318

Antibiotic exposure perturbs the gut microbiota and elevates mortality in honeybees

Kasie Raymann [1,*], Zack Shaffer [2], Nancy Moran [1]

[1]Integrative Biology, [2]University of Texas at Austin, Austin, United States

Abstract: Gut microbiomes play crucial roles in animal health, and shifts in the gut microbial community structure can have detrimental impacts on hosts. In apiculture, antibiotics are frequently used to prevent bacterial infections of larval bees, but the impact of antibiotic-induced microbial imbalance on bee health and susceptibility to disease has not been fully elucidated. Here we evaluated the effects of antibiotic exposure on the size and composition of honeybee gut communities. We monitored the survivorship of bees following antibiotic treatment in order to determine if dysbiosis of the gut microbiome impacts honeybee health, and we performed experiments to determine whether antibiotic exposure increases susceptibility to infection by opportunistic pathogens. Our results show that antibiotic treatment has persistent effects on both the size and composition of the honeybee gut microbiome. Antibiotic exposure results in decreased survivorship, both in the hive and in laboratory experiments in which bees were exposed to opportunistic bacterial pathogens. Together, these results suggest that dysbiosis resulting from antibiotic exposure affects bee health, in part due to increased susceptibility to ubiquitous opportunistic pathogens. Not only do our results highlight the importance of the gut microbiome in honeybee health, but they also provide insights into how antibiotic treatment affects microbial communities and host health.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

OT-PM4

Type VI secretion systems and associated toxins in the bee gut microbiota

Margaret Steele [1,*], Waldan Kwong [2], Eva Frederick [3], Marvin Whiteley [1], Nancy Moran [4]

[1]Molecular Biosciences, University of Texas at Austin, Austin, United States, [2]University of British Columbia, Vancouver, Canada, [3]University of Texas at Austin, Austin, United States, [4]Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: Many Gram-negative bacteria, including species found in the guts of animals, encode type VI secretion systems (T6SSs) that are used to inject protein toxins into nearby bacterial competitors. These antagonistic interactions are likely to affect the composition and stability of the microbiota, but the role of T6SSs in the evolution of gut communities is not well understood. *Snodgrassella alvi* (Neisseriales) is a specialized gut microbe of honey bees and bumble bees with high strain diversity within individual bees. Examination of the genomes of 31 *S. alvi* strains from 15 bee host species identified two conserved T6SSs and a diverse pool of T6SS-associated Rhs toxins. The architecture of the Rhs loci in *S. alvi* is consistent with the C-terminal displacement model of toxin diversification. Some of the toxins present in *S. alvi* are also found in the co-resident bacterium *Gilliamella apicola*, suggesting that toxin diversity in the bee gut is driven by both horizontal gene transfer and recombination between toxin genes. Transcriptome sequencing and re-analysis of a published transposon-mutagenesis dataset indicate that *S. alvi* T6SSs are upregulated *in vivo* and that the Rhs toxins have antibacterial activity in the bee gut. Our findings suggest that the coevolutionary dynamics of the bee gut microbiota favor both rapid toxin diversification and maintenance of T6SS machinery. T6SS-mediated competition may therefore be an important driver of coevolution within the bee gut microbiota.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

OT-PM6

Modeling human gut microbiome community structure across healthy and diseased states in 2,500 twins

Emily R. Davenport [1,*], Tim D. Spector [2], Ruth E. Ley [3], Andrew G. Clark [1]

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, United States, [2]Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom, [3]Department of Microbiome Science, Max Planck Institute for Developmental Biology, Tübingen, Germany

Abstract: Historically, bacteria were either seen as pathogenic if they caused disease or benign if they lived commensally with the host. In many cases, however, single organisms have not been identified that are consistently associated with disease, but rather it is thought that bacterial community structure and bacterial interactions differ between healthy and diseased individuals. Investigating the microbiomes of healthy and diseased individuals using systems biology methodology could lead to insight into the processes that underlie dysbiosis. To that end, methodology has been developed to identify co-occurrence networks from microbiome data, but the focus has remained on healthy microbiome datasets, and we still lack an understanding of the common properties of dysbiosis. To address these gaps, we built microbiome co-occurrence networks using 16S rRNA data from ~2,500 individuals from the United Kingdom Adult Twins Registry stratified by health status for 34 immune diseases and 133 quantitative phenotypes. First, we identified disease-associated taxa using generalized and linear mixed models that take into account relatedness between individuals. Next, we built microbiome co-occurrence networks separately for individuals i) with and without disease or ii) from opposite tails of quantitative phenotype distributions. These networks were used to identify community differences across healthy and diseased states, including comparing general network statistics (modularity and diversity), characterizing the properties of disease-associated nodes (degree, betweenness, and closeness centrality), and identifying modules of co-occurring taxa. Using these data, we are conducting one of the first large scale comparisons of microbiome dynamics across healthy and diseased individuals.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-316

The contribution of dormancy to microbial evolution

Will Shoemaker [1,*], Jay Lennon [1]

[1]Biology, Indiana University, Bloomington, United States

Abstract: Most microorganisms typically experience environmental conditions that are sub-optimal for growth and reproduction. Such conditions should influence the evolutionary dynamics of microorganisms in various ways. For example, prolonged exposure to resource-limited conditions may impose a selective pressure for increased resource use efficiency. In addition, fluctuations in resource availability can select for organisms to enter a temporary state of reduced metabolic activity (i.e. dormancy), potentially decreasing the efficiency of evolutionary processes. However, microbial evolutionary dynamics are rarely examined in resource-poor environments, as the majority of experimental evolution studies have been conducted on model organisms in environmental conditions that are optimal for growth. Here we report the results from a long-term evolution experiment of six microbial strains from different bacterial phyla, where replicate populations were transferred every 1, 10, or 100 days. Using a combination of population genetic and physiologically relevant trait data, we describe the generality of evolved responses across phylogenetically diverse taxa. From here we describe taxon-specific responses, ranging from increased resource efficiency to an increased ability to persist in a metabolically inactive state.

Expanded summary*: The goal of my research is to expand the field of experimental evolution by incorporating the relevant set of ecological and population genetic dynamics that can influence the molecular and trait evolution of microorganisms. The project described in the abstract is an attempt to understand the role of the ecologically relevant state of starvation via nutrient limitation. Using experimental evolution and a phylogenetically and metabolically diverse set of taxa where the mutation rate is known, I will describe the general and taxon-specific responses of microbial populations to extended periods of nutrient limitation. Specifically I will focus on the evolution of increased resource efficiency among taxa using carbon isotopes and the role of dormancy through the use of knockouts of the spore-forming genus *Bacillus*. I will expand on these phenotypic results be examining the change in allele frequencies over time, quantified by pooled population sequencing. Using this approach I will determine to what extent taxon-specific responses influence the molecular evolutionary dynamics of populations persisting in a resource-limited state.

I view my graduate research as an attempt to understand the role of ecological dynamics in the evolution of microorganisms through a population genetic and ecological framework. Historically, attempts to merge ecology and evolution have primarily focused on the evolution of interactions among species and individuals and their adaptive significance. While the underlying framework of evolutionary ecology has allowed researchers to test new hypothesis and develop a deeper understanding of evolution, it has often overlooked the population genetic constraints on the set of evolutionary dynamics that can occur. A primary goal of my graduate research is to integrate commonly observed ecological dynamics into the population genetics of evolving microbial populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

OT-PM3

Dynamics of intra-host symbiont evolution under mixed modes of transmission

Shelbi Russell [1,*], Russ Corbett-Detig

[1]Molecular Cellular and Developmental Biology, University of California Santa Cruz, Santa Cruz, United States

Abstract: Host-associated microorganisms exist as populations within host tissues, numbering in the millions to trillions of cells per host individual. Furthermore, symbiont populations, be it viral or bacterial, undergo multiple generations to divide and populate the host. These two attributes distinguish symbionts from their hosts and warrant investigation into the evolutionary processes that govern genetic diversity in intra-host populations. Recent research has presented evidence of mixed bacterial infections consisting of multiple strains of bacterial symbionts. Mixed infections of the chemosynthetic symbiont of the bivalve *Solemya velum* are particularly fascinating because this association undergoes vertical transmission at each host generation with occasional horizontal transmission events, which are reflected in the diversity of its intra-host populations. This observation suggests that genetic signatures of mixed infections are eroded in a few host generations. To address this question and investigate the dynamics of intra-host symbiont evolution, we developed and applied a flexible forward-time simulation framework. We find that, consistent with expectations, in the absence of horizontal transmission intra-host symbiont diversity consists of many low frequency mutations when symbionts are allowed to recombine. In the absence of recombination, we observe clonal haplotypes at correlated frequencies. When horizontal transmission is allowed, the intra-host symbiont allele frequency spectrum reflects the abundance of the novel genotype and the number of segregating sites between it and the resident genotype, consistent with observations from natural populations. The number of host generations over which mixed infection-generated diversity is retained depends on the recombination rate and bottleneck size during vertical transmission, with higher recombination rates and smaller bottleneck sizes eroding the signal more quickly. Thus, these results indicate that evidence of mixed infections is potentially ephemeral and eroded by genetic drift. Future work will address how selection acts on intra-host diversity, as novel symbiont variants introduced through horizontal transmission may experience different selective pressures than the resident symbiont.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-317

Whole-genome shotgun metagenomics highlights both intra-species co-existence and competition in the honeybee gut microbiota

Kirsten Ellegaard [1,*], Philipp Engel [1]

[1]Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland

Abstract: The mammalian gut typically harbors hundreds of bacterial species, where the taxonomic composition varies both within and between host species. Moreover, several recent studies have demonstrated a remarkably high level of intra-species diversity. However, strain-level diversity analysis in the setting of complex bacterial communities is highly challenging, as it requires extremely deep sequencing and ideally also comprehensive genome databases.
In the honeybee, the gut microbiota is dominated by only 8-10 bacterial species. Nevertheless, as for the vertebrate gut microbiota, these species are composed of a multitude of strains with highly variable gene content. Thus, the honeybee model provides a unique opportunity for studying strain-level dynamics and diversity in depth, in the setting of a natural community.
We have collected honeybees representing different colonies and life-stages, over a time-period of 2 years. Gut bacterial communities were sampled individually and subjected to whole-genome shotgun sequencing, resulting in deep coverage sequence data for each bee. Using an in-house developed pipeline, we have analyzed the distribution and extent of co-existence of closely related sub-species lineages, the gene content variation and the SNP profiles.
Our study shows that, for some species, multiple sub-species lineages co-exist in a remarkably stable manner within individual honeybees, indicative of niche-specialization or cross-feeding interactions. However, within sub-species lineages, highly divergent SNP-profiles suggest competitive exclusion, with substantial standing variation instead being maintained at the colony level. Taken together, our results underscore the need to move beyond species-level metagenomics, in order to understand the function and evolution of gut microbial communities.

Expanded summary*: The mammalian gut microbiota is typically composed of hundreds of bacterial species, where the taxonomic composition varies both within and between host species. Moreover, the complexity observed at the species level is most likely just the tip of the iceberg, as several recent studies have demonstrated a high level of intra-species diversity **(1)**. However, despite advances in sequencing technologies, strain-level analysis in the setting of complex bacterial communities represents a formidable challenge, as it requires extremely deep sequencing and ideally also comprehensive and representative genome databases.
Unlike most animals and other invertebrates, the honeybee gut is colonized by a remarkably specific and stable core bacterial community, which is dominated by only 8-10 bacterial species **(2,3)**. However, as for the vertebrate gut microbiota, the species are composed of a multitude of strains, with highly variable gene content **(4,5,6)**. These characteristics make the honeybee an attractive model for studying strain-level dynamics and diversity in depth, in the setting of a natural community.
We have collected honeybees representing different colonies and life-stages, over a time-period of 2 years, and sampled their gut bacterial communities individually. Thus, using whole-genome shotgun sequencing, we have obtained deep coverage sequence data for each bee. To take fully advantage of the honeybee model, we have established a comprehensive database, which typically allows us to map more than 90% of all bacterial-derived reads. Moreover, we have identified sub-species lineages which form discrete sequence clusters, as determined from clustering analysis of core genes derived from directly from de *novo* metagenome assemblies. Finally, using an in-house developed pipeline, we have analyzed the distribution and extent of co-existence of closely related sub-species lineages, the gene content variation and the SNP profiles.
Our study shows that, for some species, multiple sub-species lineages co-exist in a highly stable manner within individual honeybees, indicative of niche-specialization or cross-feeding interactions. However, within sub-species lineages, highly divergent SNP-profiles suggest competitive exclusion, with substantial standing variation instead being maintained at the colony level.
Taken together, our results underscore the need to move beyond species-level metagenomics, in order to understand the function and evolution of gut microbial communities.

1. Ellegaard, KM. and Engel, P. (2016): Beyond 16S rRNA profiling: intra-species diversity in the gut microbiota. *Frontiers in Microbiology*. Vol. 7: 1475

2. Engel, P. and N.A. Moran, The gut microbiota of insects - diversity in structure and function. *FEMS Microbiol Rev*, 2013. **37**: p. 699-735.
3. Sabree, Z.L., *et al*., Independent studies using deep sequencing resolve the same set of core bacterial species dominating gut communities of honey bees. *PLoS One*, 2012. **7**(7): p. e41250.
4. Engel, P., V.G. Martinson, and N.A. Moran, Functional diversity within the simple gut microbiota of the honey bee. *Proc Natl Acad Sci U S A*, 2012. **109**(27): p. 11002-7.
5. Engel, P., *et.al*, Hidden diversity in honey bee gut symbionts detected by single-cell genomics. *PLoS Genet*, 2014. **10**(9): p. e1004596.
6. Ellegaard, K.M., *et al*., Extensive intra-phylotype diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC Genomics*, 2015. **16**.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-319

Comparative soil microbiome analysis pre- and post-herbicide treatment

Chase Ridenour [1,*], Jill Cocking [1], Andrew Krohn [2], Crystal Hepp [1]

[1]School of Informatics, Computing, and Cyber Systems, [2]Biology Department, Northern Arizona University, Flagstaff, United States

Abstract: The application of broad-spectrum herbicides to crops is a common farming practice to prevent the proliferation of weeds. There have been many studies examining impacts of herbicide on the ecosystem, however, these primarily focus on the effects herbicides on plants and other macroorgansims. The few that have investigated the impacts on the microorganisms the concentration has been placed on functional or community shifts at high taxonomic levels (e.g. Phylum). Recent publications describing microbiomes of humans and other animals has emphasized the importance of using finer taxonomic resolutions to describe a system and this aspect of soil microbiomes is understudied. We compared soil samples collected from a barely field before and after the application of the broad spectrum herbicide WideMatch. 16S rRNA amplification protocol was followed per specifications of the Earth Microbiome Project on the (515/806) V4 subunits. The Quantitative Insights Into Microbial Ecology (QIIME) package was used to Qiime was used to describe the bacterial and archaeal communities of resulting amplicon pools. Comparison of weighted-Unifrac distances revealed no significant differences between treatments. However unweighted-Unifrac showed a significant ($p < 0.038$) shift in diversity between pre- and post-treatment samples. Interestingly, when considering pre- and post-herbicide application in geographically similar regions of the field, we found that soils taken in the drier portion of the field were the most compositionally different at our two sampling time points. Finally, this experiment revealed that ~ 150 OTUs and ~300 KEGG Pathways, determined with Picrust, had significantly different representation in the pre- versus post-herbicide treatment samples, including genera and KEGG Pathways associated with nitrogen fixation. These results indicate that microbial species displacement is an important consideration when characterizing impacts of herbicide on the agroecosystem.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-320

Some bee-coevolved bacterial strains are tolerant to high concentrations of glyphosate

Erick V S Motta [1,*], Nancy Moran [1]

[1]Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: Glyphosate, the primary herbicide used for weed control in agriculture, targets the EPSPS enzyme in the shikimate pathway found in plants and some microorganisms. Thus, glyphosate may affect the microbiota of animals living near agricultural sites, including pollinators such as bees. The bee gut microbiota is dominated by nine bacterial species including *Snodgrassella alvi* and *Gilliamella apicola*, which have been found to affect bee growth and pathogen susceptibility. These bacterial species exhibit a high level of strain diversity and diversification of metabolic capabilities. We identified a functional shikimate pathway in 12 *S. alvi* and 10 *G. apicola* genomes in a metagenomic screen, suggesting that these strains are susceptible to glyphosate. We therefore predict that glyphosate may alter the composition of the bacterial gut community. Some *S. alvi* and *G. apicola* strains are tolerant to high concentrations of glyphosate *in vitro*, which may be due to the presence of an insensitive EPSPS enzyme. A phylogenetic tree based on the amino acid sequence of the EPSPS enzyme clustered *S. alvi* and *G. apicola* strains with other plant and bacterial species containing glyphosate-sensitive EPSPS enzyme, suggesting that specific amino acid changes contribute to glyphosate resistance in some *S. alvi* and *G. apicola* strains. We conclude that long-term exposure of bee species to pesticides, such as glyphosate, may promote changes in their microbial community, potentially affecting bee health.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Probing Microbiome dynamics

POA-321

Rates of Metazoan Microbiome Divergence

Alex Nishida [1,*], Howard Ochman [2]

[1]Institute for Cell and Molecular Biology, [2]Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: The gut environment is subject to a constant influx of microbial colonizers, and yet, metazoan species generally harbor distinct microbiomes and can be readily differentiated based on their resident microbes. Both genetic and environmental factors, such as host phylogeny and diet, have been show to contribute to microbiome composition. To explore how species accumulate differences in microbiome composition over a broad evolutionary timescale, we analyzed the relationship between host divergence time and microbiome dissimilarity. By assembling the microbiome compositions of 150 species, we investigated how host taxonomy, diet and body mass contribute to differing rates of microbiome diversification in Mammalia and Aves. We find that host dietary changes have a significant impact on the rate of microbiome divergence.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

OTH-SI1

Evolutionary stability in fungal-bacterial symbioses

Teresa Pawlowska [1,*]

[1]School of Integrative Plant Science, Cornell University, Ithaca, United States

Abstract: Fungal-bacterial symbioses are increasingly recognized as widespread and important in medical and agronomic applications. In addition to these practical aspects, interactions of fungi with bacterial endosymbionts are a source of insights into the evolution of heritable symbioses and mechanisms that stabilize them. For example, *Ca*. Glomeribacter gigasporarum, an ancient nonessential obligate mutualist of arbuscular mycorrhizal fungi (AMF, Glomeromycota) exemplifies a nondegenerative genome reduction trajectory in heritable endobacteria. Another ancient heritable endosymbiont of AMF, *Ca*. Moeniiplasma glomeromycotorum, whose lifestyle has not been elucidated yet, exhibits extraordinary levels of intrahost diversity that can be explained by several scenarios involving different lifestyles and evolutionary histories: (i) an antagonist, (ii) a defensive mutualist, and (iii) a metabolic mutualist derived from hypervariable parasitic mycoplasmas. In contrast to symbioses of AMF that remain genetically intractable, the symbiosis between *Rhizopus microsporus* and *Burkholderia* endobacteria can be manipulated experimentally. The *Rhizopus-Burkholderia* heritable mutualism originated from an antagonism in a process likely mediated by symbiont gaining control of one of the host reproductive modes, and is offering insights into the mechanisms responsible for the host addiction to endobacteria.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-281

Population structure of a polyploid bacterial symbiont

Francine Arroyo [1,*], Esther Angert [1]

[1]Microbiology, Cornell University, Ithaca, United States

Abstract: The coral reef dwelling surgeonfish *Naso tonganus* hosts *Epulopiscium* sp. type B, an exceptionally large, highly polyploid, intestinal bacterium that reproduces only through the formation of multiple intracellular offspring. Since this symbiont is not vertically transmitted, its unique mode of reproduction and ostensibly limited access to a receptive host leads to the regular occurrence of population bottlenecks. These bottlenecks could impact the distribution and evolution of *Epulopiscium*. We investigated population dynamics of *Epulopiscium* from surgeonfish collected near Lizard Island, Australia, over the course of two decades. Multi-locus sequence analysis was used to genotype host and individual symbionts within these hosts. We found that *Epulopiscium* exhibited an extreme rate of recombination compared to point mutations (r/m = 5,344) suggesting that these symbionts rely on recombination to overcome the deleterious effects of sequential bottlenecks and clonal interference. We detected significant patterns of codivergence between the host and symbiont phylogenies but only a few populations contributed to the signal. These results suggest that *Epulopiscium* sp. type B is dependent on its host but the host may have only a facultative association with the symbiont. Although host phylogenies could not be adequately resolved to determine genetic structuring, symbiont populations showed evidence of spatial but not temporal partitioning. Surgeonfish are long-lived and likely capable of traveling long distances, yet the population structure of this prominent intestinal symbiont suggests that adult fish do not roam beyond a limited locale.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

OTH-SI4

Joining the team: Co-obligate symbionts of aphids evolving from distantly related bacteria

Alejandro Manzano-Marín [1,*], Armelle Coeur D'acier [1], Anne-Laure Clamens [1], Celine Orvain [2], Corinne Cruaud [3], Valerie Barbe [3], Emmanuelle Jousselin [1]

[1]INRA, UMR1062 (INRA, CIRAD, IRD, Montpellier SupAgro) - Centre de Biologie pour la Gestion des Populations (CBGP), Montferrier-sur-Lez, [2]CEA-Institut de Bilogie Francios Jacob-Genosope, [3]CEA-Institut de Bilogie Francois Jacob-Genoscope, Evry, France

Abstract:

Typically, aphids house the obligate nutritional bacterial symbiont *Buchnera* inside specialised cells called bacteriocytes. *Buchnera* supplies the aphid with essential amino acids and vitamins thus insuring the correct development of its host. However, some *Buchnera* lineages have lost the ability to fulfil this role, either triggered or rescued by new and younger endosymbionts. One such case are the aphid species from the Lachninae subfamily, where an ancient loss of the riboflavin biosynthetic genes in the genome of *Buchnera* was accompanied by the acquisition of a co-obligate partner, putatively *Serratia symbiotica*. However, co-obligate symbioses are not restricted to this subfamily, and examples of these have been previously reported mainly by microscopic studies. Through whole genome sequencing, we have reconstructed the genomes of the co-obligate endosymbionts from several aphid species belonging to different subfamilies, mainly the Lachninae. We have corroborated that these co-obligate symbionts indeed supplement essential metabolic auxotrophies found in *Buchnera*, mainly that of riboflavin. Not surprisingly, they have evolved genomes with similar core metabolic capabilities, with some even bringing new ones to the symbiotic consortium. Also, we have determined that these co-obligate symbionts have evolved from diverse facultative symbiotic taxa associated to aphids as well as free-living bacterial strains. These findings show that co-obligate symbiosis in aphids is more widespread than previously thought. This suggests a fragile mono-symbiotic association between the aphid host and its *Buchnera* symbiont, whose highly degenerated genome could undergo simple metabolic losses which could lead to a "lucky" secondary symbiont establishing as a co-obligate one.

Expanded summary*: Investigating the rise of obligate symbioses across aphids (Hemiptera: Aphididae)

Mutualistic symbiotic relationships between prokaryotic microorganisms and eukaryotic hosts are widespread in nature and are suspected to have played a key role in the diversification of these. These symbiotic associations endow the eukaryotic host with an expanded set of metabolic capabilities and adaptive traits . Within insects, mutualistic symbiotic associations with prokaryotic microorganisms are prevalent. These associations enable different insects to feed on nutritionally-unbalanced diets, such as blood and plant sap. Among insects, the best-described animal-microbe symbiosis is that of the pea aphid (Hemiptera: Aphididae) and its obligate nutritional mutualist, *Buchnera aphidicola* (Gammaproteobacteria). This symbiont resides in specialised cells, called bacteriocytes, and provides the aphid with essential amino acids and some vitamins (namely biotin and riboflavin) that are lacking in their strict phloem diet.

Besides *Buchnera*, aphids can harbour a variety of secondary symbionts. These secondary symbionts can be of facultative or obligate nature. Contrary to obligate symbionts, facultative ones are not required for the correct development, reproduction and survival of their host, although they can provide an advantageous trait under certain environmental or ecological conditions. The Lachninae subfamily of aphids is particular, having all analysed members to date displaying apparent secondary obligate endosymbionts. The genome sequencing of three endosymbiotic pairs within aphids belonging to this subfamily has revealed an ancient loss of the riboflavin (vitamin B_2) and some biotin (vitamin B_7) biosynthetic genes in *Buchnera* from the Lachninae last common ancestor (LLCA). This loss would have triggered the establishment of a secondary co-obligate endosymbiont.

However, the phenomenon of co-obligate symbiosis has been described in aphids species within different subfamilies. Thus, through the massive sequencing of different aphids across the phylogeny we intend to explore the bases for these suspected co-obligate symbiosis as well as the identity of the bacterial symbiotic partners. First, we have discovered that these co-obligate symbionts have actually arose from several taxa: e.g. *Wolbachia*, *Serratia symbiotica*, *Erwinia*, and *Sodalis*. Through the genome-based metabolic annotation of these secodnary symbionts, we have determined that these are capable of rescuing *Buchnera*'s deficiencies in producing

essential nutrients. Nonetheless, we have found evidence that the genes involved in these vitamin-biosynthetic pathways have sometimes been horizontally-acquired by the secondary endosymbionts from different sources. Additionally, the phylogenetic reconstruction of the aphids relationships through the use of both *Buchnera* and mitochondrial genes is aiding us in understanding the number of times these co-obligate associations have arose as well as the order of symbiont succession.
These project forms part of a larger project exploring symbiont replacement in different aphids, particularly within the *Cinara* genus, representing more than half of the Lachninae subfamily diversity. These findings will reveal fundamental principles in obligate symbiont settlement (free-living -> facultative -> obligate mutualistic) and will shed light to the different adaptations a free-living or facultative symbiont undergoes when changing its lifestyle. Furthermore, given that, at least within *Cinara*, some symbiont shifts coincide with shifts in host-plant taxon, we expect these project will give us clues, if any, into the role these symbiont shifts have played in the diversification of their hosts.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-279

Comparative genomics getting at the roots of a dual symbiosis: the importance of HGT

Amanda Brown [1,*], Dana Howe [2], Sulochana Wasala [2], Amy Peetz [3], Inga Zasada [3], Dee Denver [2]

[1]Texas Tech University, Lubbock, [2]Oregon State University, [3]USDA-ARS Horticultural Crops Research Laboratory, Corvallis, United States

Abstract: A major question in symbiosis research is how parasitic or free-living organisms transition to stable beneficial relationships with their hosts. Some endosymbionts go all the way to become obligate partners, while most seem to persist in facultative roles from the host's perspective. When multiple endosymbionts persist together theory predicts they would encounter the negative effects of potential niche overlap and competition. This could drive them to specialize in different tissues, cells, or form different relationships with their hosts. We looked at a puzzling dual-symbiosis with two of the most widespread terrestrial endosymbionts, *Wolbachia* (Alphaproteobacteria) and *Cardinium* (Bacteroidetes), which are extensively found in insects and spiders/mites, respectively. In these hosts, these unrelated symbionts can co-occur and have converged in many ways, including independently evolving the same host-manipulating phenotype, cytoplasmic incompatibility (CI). We focused on this dual-symbiosis discovered in the root lesion nematode, *Pratylenchus penetrans*. We performed genome sequencing, phylogenomics, comparative genomics, and fluorescence in situ hybridization (FISH). Phylogenetic analysis showed that both *Wolbachia* and *Cardinium* have long histories within plant parasitic nematodes, likely being the earliest host association for *Wolbachia* and second earliest for *Cardinium*. Based on signs of comparative genome streamlining (genome size, gene density, protein length, and rates of repetitive/parasitic elements such as transposons), this *Wolbachia* strain was more streamlined than the *Cardinium* strain (e.g. *Wolbachia*: genome size 975,127 bp, 962 proteins, 32.1%G+C, 86.6% coding versus *Cardinium:* genome size 1,358,214 bp, 1,131 proteins, 35.8% G+C, 79% coding), suggesting a tighter host association. The *Cardinium* strain also showed less biosynthetic capacity than *Wolbachia*, suggesting more dependence on the host, or perhaps even some dependence on *Wolbachia*. By contrast, other *Wolbachia* and *Cardinium* strains emerging later in their respective clades showed additional biosynthetic capacity through probable horizontal gene transfer (HGT). Remarkably, as shown in previous studies, these ancient gene transfers appear to involve exchanges between these two dual-symbionts, and may have been key events on the path to obligate mutualism. Our data from *P. penetrans* showed this *Cardinium* genome encoded many genes probably originating from horizontal transfer from alphaproteobacteria (including *Wolbachia* and *Rickettsia*), and several genes potentially originating from the host nematode. However, we found no genes homologous to CI candidate genes in either genome, while FISH results, alliance with mutualist plant sap feeders, and genome repertoire hinted at one or both endosymbionts having a possible nutritional symbiotic role. In conclusion, we hypothesize that ongoing lateral gene transfer may be important in the early history and present evolutionary trajectory of these successful symbionts, consistent the hypothesis of dual endosymbiosis providing a "training grounds", or considering HGT, a "trading floor" for future mutualism.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-294

Comparative transcriptomics and phylogenomics of the evolution of nitrogen-fixing symbiosis in Betulaceae

Sitaram Rajaraman [1,*], Olli-Pekka Smolander [2], Gugan Eswaran [1 2], Airi Lamminmäki [1], Ari-Pekka Mähönen [1 2], Petri Auvinen [2], Jarkko Salojärvi [1]

[1]Department of Biosciences, [2]Institute of Biotechnology, University of Helsinki, Helsinki, Finland

Abstract:

Family *Betulaceae* is in an interesting phylogenetic position for addressing evolution of nitrogen fixing symbiosis and host-symbiont interaction. The family contains morphologically and genetically very similar pioneer tree species, birches and alders, which have evolved different strategies for nitrogen fixation. Birches, for example *Betula pendula*, form ectomycorrhizal associations with fungi (such as *Paxillus involutus*) for nitrogen uptake, while alders, such as *Alnus glutinosa*, are a nodulating species, forming mutualistic associations with bacterium *Frankia alni*.

We study the evolution of differential symbiotic states of *Alnus* and *Betula* through a comparative genomics and transcriptomics approach. In addition to the recently submitted *B.pendula* genome, we have recently carried out a full PacBio assembly of *A.glutinosa*, and genome annotation is currently on-going. Comparative genomics analysis will address the genetic constraints controlling the regulation, localisation and functional adaptation of genes, pathways and host-symbiont recognition mechanisms. Early results indicate the conservation of the full nod-signalling pathway in *B.pendula*.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-295

Defining Brugia malayi and Wolbachia symbiosis by stage-specific Dual RNA-seq

Alexandra Grote [1,*], Tao Ding [1], Nirvana Nursimulu [2], Sara Lustigman, John Parkinson [2], Elodie Ghedin [1]

[1]Biology, New York University, New York, United States, [2]University of Toronto, Toronto, Canada

Abstract: Filarial nematodes represent one of the leading causes of disability in the developing world. Many filarial worm species, including *Brugia malayi*, one of the causative agents of lymphatic filariasis, have an obligate endosymbiotic relationship with the alpha-proteobacteria *Wolbachia*. Elimination of the endosymbiont interferes with development, reproduction, and survival of the worms within the host, a clear indicator that *Wolbachia* are crucial for survival of the parasite, yet little is understood about the mechanism underlying this symbiosis. To better understand the molecular interplay between these two organisms we profiled the transcriptomes of *B. malayi* and *Wolbachia* by dual RNA-seq across the life cycle of the parasite. This helped us identify the functional pathways involved in this essential symbiotic relationship provided by the co-expression of nematode and bacterial genes. We have identified significant stage-specific and gender-specific differential expression in *Wolbachia* during development of the worm. For example, during female worm development we find that *Wolbachia* upregulate genes involved in ATP production and purine biosynthesis, as well as genes involved in the oxidative stress response. We are currently using these data to characterize the endosymbioic relationship at the metabolic level using Flux Balance Analysis, identifying choke points that could be exploited for therapy. We used metabolic reconstruction to create a draft metabolic network with available enzyme annotation data for the genomes of *B. malayi* and *Wolbachia.* Using *in silico* knockouts, we will determine the necessary pathways for growth and virulence and determine how these pathways are influenced by the presence of *Wolbachia.*

Expanded summary*: Human filarial infections are currently a leading cause of morbidity in the developing world. Despite the large cost to human health, the chronic and debilitating diseases caused by filarial nematodes remain largely neglected. Two of the most prevalent chronic diseases due to filarial nematodes include lymphatic filariasis, caused by *Wuchereria bancrofti*, *Brugia malayi*, and *Brugia timori*, and onchocerciasis, caused by *Onchocerca volvulus*. Currently 38.5 million people have lymphatic filariasis, representing over 200,000 years lived with disability (YLDs) in 2015, while 15.5 million people have onchocerciasis, representing over 100,000 YLDs in 2015. While efforts to mitigate the effects of these diseases have been successful in some regions, current medications are insufficient to reach elimination by 2020, particularly in regions of co-endemicity with loasis, caused by the filarial nematode *Loa loa*. Current mass drug administration relies on a small arsenal of drugs, increasing the likelihood of development of resistance, a phenomenon already observed in their veterinary applications.

Most filarial nematodes have an obligate endosymbiotic relationship with intracellular bacteria of the genus *Wolbachia*. As the filaria require these bacteria to develop, reproduce and survive in the human host, they represent an attractive target for intervention. While the relationship between the nematode and the bacteria is known to be co-dependent, the molecular basis for this relationship remains poorly understood. To better understand the molecular interplay between these two organisms we profiled the transcriptomes of *B. malayi* and *Wolbachia* by dual RNA-seq across the life cycle of the parasite. This helped us identify the functional pathways involved in this essential symbiotic relationship provided by the co-expression of nematode and bacterial genes. We have identified significant stage-specific and gender-specific differential expression in *Wolbachia* during development of the worm. For example, during female worm development we find that *Wolbachia* upregulate genes involved in ATP production and purine biosynthesis, as well as genes involved in the oxidative stress response. We have also identified co-expressed pathways required for molting of the worm from L3 to L4, the molt which marks the establishment of infection in the human host. This global transcriptional analysis has highlighted specific pathways to which both *Wolbachia* and *B. malayi* contribute concurrently over the life cycle of the parasite, paving the way for the development of novel intervention strategies.

We are currently using these data to characterize the endosymbioic relationship at the metabolic level using Flux Balance Analysis to identify choke points that could be exploited for therapy. We used metabolic reconstruction to create a draft metabolic network based on available enzyme annotation data for the genomes of *B. malayi* and *Wolbachia*, as well as the *Wolbachia*-free filarial parasite, *Loa loa*. Using *in silico* knockouts, we will determine the necessary pathways for growth and virulence of the filaria and determine how these pathways are influenced by the presence of *Wolbachia*. This method will allow for the identification of drug targets specific to

B. malayi that will not affect *L. loa*, which are greatly needed due to the risk of severe adverse effects in the treatment of co-infected patients.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-276

Drivers of the rapid evolution of bag of marbles in Drosophila: evaluating functional diversification and a conflict with Wolbachia

Jaclyn Bubnell [1,*], Charles Aquadro [1]

[1]Molecular Biology and Genetics, Cornell University, Ithaca, United States

Abstract: In *D. melanogaster*, *bag of marbles* (*bam*) acts as the master switch for germline stem cell differentiation during oogenesis in females and plays a key role in regulating normal spermatogenesis in males. Interestingly, *bam* is rapidly evolving across the *Drosophila* genus and shows a signal of positive selection in both the *D. melanogaster* and *D. simulans* lineages, but not in several other species. What could be driving positive selection of a gene essential for fertility? One possibility is that *bam's* function as the master switch for oogenesis has arisen in the lineage leading to *D. melanogaster* and *D. simulans* and now this novel function is under positive selection. To evaluate this hypothesis, we are using CRISPR-Cas9 to generate *bam* null alleles in diverse *Drosophila* species to test for conservation of function. Our results in *D. yakuba* show that *bam* null females have similar phenotypes to *D. melanogaster bam* null females. In both species, oogenesis is completely disrupted as females are sterile and contain large tumorous ovaries. These data suggest that *bam* is also regulating oogenesis in *D. yakuba*. Preliminary population genetic data in *D. yakuba* indicates that in contrast to *D. melanogaster* and *D. simulans, bam* is not evolving under positive selection. Together, these results imply that although *bam* shows different evolutionary patterns across these species, its core function is conserved, and therefore a novel function may not be driving positive selection in *D. melanogaster* and *D. simulans.* If novel *bam* function is not driving positive selection, another possible driver is genetic conflict with germline parasites. When the *D. simulans bam* sequence is expressed in *D. melanogaster,* female fertility and germline stem cell regulation is disrupted, but there is no affect on males. We have also observed that the maternally inherited intracellular bacteria *Wolbachia* rescues the female fertility defect caused by a *bam* hypomorphic mutant in *D. melanogaster.* Episodic natural infections with *Wolbachia* could result in a genetic conflict with *bam* and lead to an evolutionary arms race for control of oogenesis. We are currently working to better define the interaction between *D. melanogaster bam* and *Wolbachia* as well as generate *bam* hypomorphs in *D. simulans, D. yakuba,* and *D. ananassae.*

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

OTH-SI5

The symbiotic potential of Paraburkholderia bacteria in the social amoeba Dictyostelium discoideum

Tamara Haselkorn [1,*], Susanne DiSalvo [2], Suegene Noh [3], Debra Brock [3], Usman Bashir [3], Joan Strassmann [3], David Queller [3]

[1]Department of Biology, University of Central Arkansas, Conway, [2]Department of Biology, Southern Illinois University, Edwardsville, Edwardsville, [3]Department of Biology, Washington University in St. Louis, St. Louis, United States

Abstract: The ability to interrogate host-microbe interactions is what makes the amoeba, *Dictyostelium discoideum* an ideal model for studying the dynamics of symbiosis. This soil-dwelling amoeba has diverse interactions with bacteria: bacteria are prey, pathogens, and symbionts. Certain species of *Paraburkholderia* (formerly *Burkholderia*) are able to survive amoebal digestion and maintain a persistent relationship, conferring to its amoeba host the ability to carry other food bacteria within its spores, with the potential benefit of seeding its own food crop. To characterize the prevalence and diversity of this association in nature we screened over 700 isolates of *D. discoideum* and found a 25% infection prevalence. Using a multilocus phylogenetic analysis, we discovered at least two introductions of the *Paraburkholderia* symbiont into *D. discoideum*. There are at least 10 distinct symbiont strains that fall into two distinct clades, one of which shows evidence of rapid sequence evolution. To assess the symbiotic potential of these bacteria, we infected each of these symbiont strains, as well as non-symbiont *Paraburkholderia* strains into novel *D. discoideum* hosts. For the different *Paraburkholderia* strains, we found variation in their ability to colonize, persist, and enter into spores, suggesting that *Paraburkholderia*'s relationship with *D. discoideum* exists on a symbiotic continuum that will be useful for exploring the transition to the symbiotic lifestyle and mechanisms of this interaction. Only *D. discoideum Paraburkholderia* symbionts (from both clades) were able to fully express all symbiont traits, and we have sequenced the genomes of 12 symbiont and closely related non-symbiont strains to look for patterns of genome evolution consistent with the transition to the intracellular lifestyle.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-286

Mitochondrial-Y Interactions Determine Gene-Expression Profiles in Drosophila melanogaster males

Manisha Munasinghe [1,*], J. Arvid Ågren [2], Angela Early [2], Andrew Clark [1 2]

[1]Biological Statistics and Computational Biology, [2]Molecular Biology and Genetics, Cornell University, Ithaca, United States

Abstract: The presence of two genomes in animal cells is the consequence of an ancient mutualistic partnership between an archaeon and a bacterial endosymbiont. Coordination of gene expression between the mitochondrial and nuclear genome is therefore essential for the maintenance of mitochondrial function and organismal fitness. Both genomes must co-evolve in response to new mutations and this co-evolution can take place through many different avenues. For example, one consequence of the maternal inheritance of the mitochondrial genome is the potential accumulation of male-harming mutations (the Mother's Curse), which in turn creates a selective pressure in males for compensatory nuclear mutations. Theory and simulations show that we should expect an enrichment of such mutations on the Y-chromosome, due in part to their escape from female counter-selection. To explore the extent to which mito-Y combinations influence genome-wide gene expression, we performed 3'-end RNA-sequencing on 36 Drosophila melanogaster lines derived from all possible crosses of females from 6 lines with deeply divergent mitochondrial clades and 6 male lines with Y chromosomes of remote geographic origin (all having identical nuclear backgrounds). We identified an abundance of differentially expressed transcripts among the progeny. Both the Y chromosome and the mitochondrial haplotype appeared to have trans-acting effects on expression levels of many genes. While there were instances of Y-mitochondrial interaction, experiments to assess fitness effects are still underway to discern whether these are clear examples of variants that act as compensation for Mother's Curse.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-285

Divergence of bacterial gut symbionts between two floricolous drosophilids, Colocasiomyia alocasiae and C. xenalocasiae

Jia-Syuan Chen [1,*], Chau-Ti Ting [2], Shu Fang [3], Shun-Chern Tsaur [4]

[1]Department of Life Science, NTU, [2]Life Science, Department of Life Science, [3]Biodiversity Research Center, Academia Sinica, [4]Office of International Affairs, NTU, Taipei City, Taiwan

Abstract: Gut bacterial community composition and abundance profiles are shaped by host genetics, geography, evolutionary history, and also diet, bringing about coevolution between bacteria and animal hosts. Previous studies have been focused on the differences of gut bacterial community among different animal hosts. However, the divergence and the functions of gut bacterial symbionts in specialist species, which can only thrive with a limited diet, are largely unknown. Furthermore, the differences of bacterial communities between animals and diets in the nature are unclear. Here we studied the bacterial communities in the guts of two *Colocasiomyia* spp. and their Araceae host plants. *Colocasiomyia* spp. are specialists, highly associated with the inflorescences of Araceae. *Colocasiomyia alocasiae* and *C. xenalocasiae*, both feeding on the inflorescences of *Alocasia odora* and *Colocasia formosana*. We sampled and sequenced bacterial community from fly gut and on the surface of two plants from 2011 to 2016. The results showed that the bacterial flora in both sites were restricted in only few dominant bacterial families, including *Bartonellaceae, Enterobacteriaceae, Enterococcaceae*, and *Lactobacillaceae*. Although the similar bacterial taxa were observed, the composition are different. Principle coordinates analysis clustered the bacterial communities into three groups, *A. odora* feeders, *C. formosana* feeders, and plants. The two plants surprisingly had similar bacterial communities on the surfaces. The results support the diet is the primary factor shaping the gut bacterial community, the internal conditions of fly also appear to exert some level of control over the bacteria in shaping the bacterial communities in the digestive tract.

Expanded summary*: A specialist species can only be prospered in a narrow range of environmental conditions or utilize limited diet sources, compared to a generalist species, which can adapt to wide range of environment and has varied diet sources. The ecological specialization causes specialists highly relay on the specific diets. Thus, the dispersion of specialists would be limited by their diets distribution. In fact, the diet shifting and adaption are understood as the process of speciation. The host specialization occurs from a specialist species starts to change the diet source, isolating from the original population. The bacterial community in gut of a specialist species would be the first to react the changes. By changing the composition of microbiota in gut, the function of bacterial community has the plasticity to help animals adapt to the diets. Despite diet adaption is commonly in the wild, the ecological and genetic mechanisms are still largely unknown. We studied the bacterial communities in the two specialist species, *Colocasiomyia* spp. and their Araceae host plants. We observed the feeding preference of the specialists and reported the diet associated bacteria in this study, implying bacteria play some roles in diet adaption.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-427

EVOLUTION OF GENE FAMILIES IN APHID-ENDOSYMBIONT INTERACTIONS

Serena Zhao*, Nancy Moran [1]

[1]University Of Texas At Austin, Austin, United States

Poster: Microbial endosymbionts are widespread across insects. Regulation of symbiont populations is crucial to host fitness and the evolutionary trajectory of the interaction. Symbiotic function is shaped by evolution in the gene families active in the interactions, and these genes are subject to selection pressure from the symbiont population as well as from stochastic processes in the host population. Traits involved in interactions with symbionts can vary among hosts due to duplication and loss of genes. Codiversified lineages of hosts and endosymbionts provide a comparative framework for studying the evolution of gene families active in symbiotic interactions.

The obligate bacterial endosymbiont *Buchnera aphidicola* is present in almost all aphid species, having codiversified with the host following colonization of the aphid ancestor. Buchnera provides amino acids from within specialized host cells, bacteriocytes. Some peptides have been found to be highly upregulated in bacteriocytes of the pea aphid (*Acyrthosiphon pisum*), suggesting potential mechanisms of host regulation of symbiont populations. Although similar mechanisms of symbiont regulation may be at work in different aphid species, the evolutionary history of these gene families is not known.

Recently, genome assemblies for four aphid species in the subfamily Aphidinae have become available: *Acyrthosiphon pisum*, *Myzus persicae*, *Diuraphis noxia*, and *Aphis glycines*. Through comparative genomics, we aim to characterize the evolutionary history of peptides shown to be overexpressed in bacteriocytes in *A. pisum*. Duplications in these gene families predate speciation, suggesting that these genes have played a role in regulating symbionts throughout aphid evolution. Current variation in symbiont regulation due to subsequent deletions, duplications, and divergences of gene copies likely results from divergent selection pressure or stochastic conditions among the host lineages. Patterns in these genes underlying symbiont regulation will illustrate the impacts of obligate symbiosis on gene family evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-283

Population genetics applied to 200 genomes of Rhizobium leguminosarum biovar trifolii

Maria Izabel A. Cavassim [1,*], Asger Bachmann [1], Sara Moeskjær [2], Bjarni Vilhjalmsson [1], Luc Janss [2], J. P. W. Young [3], Mikkel Schierup [1], Stig U. Andersen [2]

[1]Department of Computer Science, [2]Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark, [3]Department of Biology, University of York, York, United Kingdom

Abstract: Bacteria of the genus *Rhizobium* play an important role in the sustainability of agriculture by increasing the nitrogen intake of legumes without the introduction of N-fertilizers. The symbiosis between rhizobia and legumes relies on the existence of specific bacterially encoded genes (Nif and Nod). These, and other genes, may be acquired via horizontal gene transfer (HGF) from related species, leading to an accretion of genetic diversity in a single population. The specific set of core and accessory genes maintained in a population could provide direct evidence of the evolutionary forces driving speciation.
To investigate this, we examined 200 genomes of the clover symbiont *Rhizobium leguminosarum* from European populations, looking at nucleotide diversity, site frequency spectrum, linkage disequilibrium (LD) and GC content.
Even though we sampled 200 strains from a single named species, pairwise nucleotide differences revealed the existence of 5 distinct biological species (genospecies). The median GC content of accessory genes was lower than that of core genes, as commonly observed for horizontally transferred genes.
At the level of genomic configuration, all isolates shared at least a chromosome and 2 chromids but there was great variation on the number of extra replicons. When zooming into the gene level, we observed that core genes were mostly placed on the three common replicons and that private genes, found uniquely in a genospecie, were present in most of the members. Finally, the distribution of allele frequencies and decay of intragenic LD were very different between genospecies, leading us to believe that selection and recombination is differentially shaping their nucleotide diversity.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-291

Diverse patterns of genome evolution within a genus of facultative insect-symbiotic bacteria

Wen-Sui Lo [1], Chih-Horng Kuo [1,*]

[1]Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

Abstract: The genus *Spiroplasma* contains a group of helical, motile, and wall-less bacteria in the class Mollicutes. Ecologically, most of the *Spiroplasma* species appear to be harmless commensals of insects, while a small number of species have evolved pathogenicity toward their arthropod or plant hosts. Phylogenetically, *Spiroplasma* are closely related to the animal pathogens in the genus *Mycoplasma*. However, in contrast to the extensive efforts devoted to *Mycoplasma* research, genomic studies concerning *Spiroplasma* have received much less attention. This bias in taxon sampling of available genome sequences has limited our understanding of genome evolution in these bacteria, particularly considering the observation that *Mycoplasma* is a polyphyletic group with the *M. mycoides* clade nested within *Spiroplasma*. To improve the understanding of *Spiroplasma* genome diversity and to facilitate comparative analyses with *Mycoplasma*, our research group has conducted multiple genome sequencing projects for the key *Spiroplasma* lineages. Our comparative analyses demonstrated that the patterns of genome evolution varied widely within this genus of symbiotic bacteria. For example, while the species belonging to the Citri clade all experienced extensive viral invasion into their chromosomes, which promoted rampant horizontal gene transfers and genome rearrangements, the species belonging to the sister Chrysopicola clade have maintained stable chromosomal organization since their divergence. Additionally, based on comparisons among four divergent mosquito-associated *Spiroplasma* species in the Apis clade, we found that these bacteria have adapted to similar hosts through different strategies of nutrient utilization and the horizontal transfer of few virulence genes have contributed to the evolution of pathogenicity. In summary, the diversity of *Spiroplasma*-host interactions provides a valuable system to investigate the evolutionary transitions between ecological niches in Mollicutes. Future comparative genomics and molecular genetics studies will further improve our knowledge of the genetic mechanisms that explain the phenotypic differences among species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-289

Can selective advantages explain the AT-bias of endosymbiotic genomes?

Anne-Kathrin Dietel [1,*], Christian Kost [2], Martin Kaltenpoth [3]

[1]Experimental Ecology and Evolution, Max Planck Institute for Chemical Ecology, Jena, [2]Department of Ecology and Evolution, University of Osnabrück, Osnabrück, [3]Department of Ecology, Johannes Gutenberg-University, Mainz, Germany

Abstract: The genomic base composition of bacteria is highly variable, ranging from less than 15% to more than 75% GC. While increased GC-contents occur predominantly in free-living taxa, genomes with low GC-contents are largely restricted to intracellular bacteria. Interestingly, bacterial endosymbionts, plasmids, and viruses generally exhibit increased AT-contents relative to their hosts' DNA, suggesting this pattern characterizes host-dependent elements. The evolutionary factors causing this biased base composition, however, remain elusive. Given that A is the most abundant nucleotide within living cells and A+T nucleotides are energetically less expensive than G+C, we hypothesized that extrachromosomal elements, whose fitness is limited by the availability of A + T/U, should be selectively favoured over elements requiring G + C. To test this hypothesis, we experimentally manipulated the GC-content of *Escherichia coli* cells by introducing AT- or GC-rich plasmids, respectively. Indeed, competition experiments revealed that increased GC-contents caused a drastic decrease of the cells' fitness. Furthermore, determining the plasmid copy number via quantitative real-time-PCR revealed lower copy numbers of all GC-rich plasmids compared to AT-rich plasmids, which is in line with the idea that the availability of G+C nucelotides limits replication of both plasmids and cells. Strikingly, providing GC-nucleotides to cells with higher demands for GC-nucleotides increased their fitness, while AT-supplementation did not have this effect, thus further corroborating that the availability of GC-nucleotides limited their growth. Taken together, our results provide strong experimental evidence that the commonly observed increased AT-content of host-dependent elements can be selectively favoured.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

OTH-SI2

Rates of Organelle Genome Evolution and their Effects on Cytonuclear Interactions

Daniel Sloan [1,*]

[1]Department of Biology, Colorado State University, Fort Collins, United States

Abstract: One of the defining features of the eukaryotic cell is the presence of multiple genomes (nuclear, mitochondrial, and plastid), reflecting the endosymbiotic origins of eukaryotes. Key biological functions require that the gene products encoded by these genomes interact in an intimate and coordinated fashion. Therefore, nuclear and cytoplasmic genomes must coevolve in spite of fundamental differences in how they are replicated, inherited and expressed. By taking advantage of an angiosperm genus (Silene) that exhibits extreme variation in rates of mitochondrial and plastid genome evolution, we have investigated how changes in organelle genomes can select for coevolutionary responses in the nucleus. A combination of phylogenetic analysis of evolutionary rate correlation, population genetic tests for selection, and in silico modeling of protein structures has yielded a robust picture of positive selection driven by cytonuclear coevolution. We have detected likely signatures of both mutualistic and antagonistic dynamics, reflecting the complex mixture of evolutionary forces that shape symbiotic systems.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-293

Experimental evolution of the intracellular symbiont Wolbachia identifies loci correlated with adaptation to the host

Irene Newton [1,*], Kathy Sheehan [2], Maryann Martin [1], William Nawrocki [3]

[1]Indiana University, Bloomington, IN, USA, [2]Department of Biology, [3]Indiana University, Bloomington, United States

Abstract: Many bacteria live as obligate intracellular symbionts of eukaryotic hosts and spend their entire life cycle – all generations of division – within the host environment. Because the intracellular symbionts cannot exist outside of the host, when challenged with infection of a new host, the bacteria must either colonize the new host or be purged. Here we leveraged the power of Drosophila genetics to identify how an intracellular bacterium (Wolbachia pipientis) evolves in a new host context. Wolbachia are obligate intracellular alpha-proteobacteria that infect nearly half of the insect species on the planet, making them the most prevalent bacterial infection on Earth. We discovered a Drosophila mutant that is Wolbachia deficient – that is, wild-type Wolbachia cannot colonize it. We used this extreme host bottleneck to select for Wolbachia variants, through sequential passage, that can recolonize the previously un-permissive host background, based on both western blots and quantitative PCR. We resequenced the genomes of the Wolbachia variant and identified multiple mutations, in functional categories such as secretion machinery and ankyrin repeat proteins, associated with the colonization of the mutant host background. The data suggest the mutations may be involved in interactions with the fly. Our results confirm that Wolbachia can quickly adapt to a new host background and suggest that mutator variants may have a competitive advantage, even intracellular symbiont lineages.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-292

An intracellular symbiont manipulates its host through multiple candidate effectors, correlated in their evolution and expression profile

Irene Newton [1,*], Kathy Sheehan [2], Danny Rice [3]

[1]Indiana University, Bloomington, IN, USA, [2]Department of Biology, [3]Indiana University, Bloomington, United States

Abstract: *Wolbachia pipientis* is an intracellular symbiont of arthropods well known for the reproductive manipulations induced in the host and more recently, for the ability of *Wolbachia* to block virus replication in insect vectors. Due to the fact that *Wolbachia* cannot be genetically manipulated, and inherent limitations of working with an intracellular symbiont, very little is known about mechanisms used by *Wolbachia* for host interaction. Here we employed a bioinformatics pipeline and identified 163 candidate effectors, potentially secreted by *Wolbachia* into the host cell. A total of 84 of these candidates were then subjected to a screen of growth defects induced in yeast upon heterologous expression, identifying 15 top candidates, likely secreted by *Wolbachia*. These candidate secreted effectors likely function in concert as we find that their native expression is correlated and is highly upregulated during *Drosophila* pupation. In addition, the evolutionary history of these candidates is similarly correlated, suggesting they may function together, or in the same pathway, during host infection. Interestingly, many candidates are *Wolbachia* clade specific – only found within one or two *Wolbachia* clades – perhaps reflecting shared evolutionary history or similar mechanisms of host manipulation. Finally, we use this set of identified candidates to identify a putative bioinformatic signature for *Wolbachia* type IV effectors, paving the way for future work on the mechanism of secretion in *Wolbachia.*

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-288

Evolution of tick-associated pathogens and endosymbionts

Rahul Raghavan [1,*]

[1]Biology, Portland State University, Portland, United States

Abstract:

Ticks (order Ixodida) rely solely on blood for nutrition, but blood is a poor source for several amino acids and cofactors required for normal development. Similar to insects that thrive on unbalanced plant diets, ticks depend on intracellular symbiotic bacteria such as Coxiella-like and Francisella-like endosymbionts (CLE and FLE, respectively) that likely provision nutrients missing from their diets. In addition to containing endosymbionts, ticks often vector mammalian pathogens, including Francisella tularensis and Coxiella burnetii that are closely related to FLEs and CLEs, respectively. However, the evolutionary relationship between tick-associated endosymbionts and mammalian pathogens are not clearly understood. In order to gain new insights into the evolution of CLEs and FLEs, we sequenced and analyzed the genomes of the CLE present in Amblyomma americanum (Lone star tick) and the FLEs present in Amblyomma maculatum (Gulf coast tick) and Ornithodoros moubata (African relapsing fever tick). Our data show that while pathogenic Coxiella is a sister taxon of CLEs, FLEs were derived from pathogenic Francisella. In addition, our analyses indicate that the FLEs are more recent endosymbionts than CLEs and that multiple acquisitions of FLEs have occurred in soft-backed ticks.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-287

Host response to increased symbiont complexity via increased symbiont load

Matthew Campbell [1,*], Piotr Lukasik [1], John McCutcheon [1]

[1]DBS, University of Montana, Missoula, United States

Abstract: Approximately two billion years ago, a bacterium entered into endosymbiosis with an archaeal host cell and became the mitochondrion. Some bacterial endosymbionts from sap-feeding insects are strikingly similar to organelles, and so might be used as models to study early organelle evolution. Previously we have shown that an endosymbiont of cicadas, *Candidatus* Hodgkinia cicadicola (*Hodgkinia*), has split into dozens of cytologically distinct genomic lineages in the periodical cicada *Magicicada tredecim.* Here we show this endosymbiont complexity exists in all seven *Magicicada* species, in each case resulting in essential genes being subdivided onto numerous genomes present at very different abundances. We have studied the consequences that spitting has on transovarial transmission of *Hodgkinia* from parent to offspring. With increased *Hodgkinia* complexity, the host cicada must either transmit more *Hodgkinia* cells to each egg to ensure each egg gets the full complement of *Hodgkinia* genes, or lay a large proportion of inviable eggs that do not receive all *Hodgkinia* genomes. Using 16S rRNA amplicon sequencing and microscopy, we show that *Magicicada* transmit a larger number of *Hodgkinia* cells to each egg compared when to cicadas with less complex *Hodgkinia*. While this ensures eggs contain all *Hodgkinia* lineages, we hypothesize that this increased endosymbiont cell load imposes a cost on the cicada host and reflects the nonadaptive nature of *Hodgkinia* genome splitting.

Expanded summary*: Approximately two billion years ago, a bacterium entered into endosymbiosis with an archaeal host cell and became the mitochondrion. Early in its evolution, the new organelle underwent dramatic gene loss and integration with the host genome. Sap feeding insects harbor additional bacterial endosymbionts that provide nutrients to supplement the host's nutrient-poor diet. These endosymbionts are strikingly similar to organelles – they are inderdependent with their hosts, have reduced genomes, and sometimes require nuclear encoded genes for proper function – so therefore might reveal something about how separate organisms become interdependent genetic and cell biological entities.

Previously we reported that an endosymbiont of cicadas, *Candidatus* Hodgkinia cicadicola (*Hodgkinia*), has split into dozens of cytologically distinct genomic lineages in the periodical cicada *Magicicada tredecim.* Here we show this complexity exists in all seven *Magicicada* species as a result of an ongoing lineage splitting process. Because this genomic splitting also results in *Hodgkinia* genes being distributed among numerous genomes present at very different abundances, it has consequences for the transovarial transmission of *Hodgkinia* cells to cicada eggs. With increased *Hodgkinia* complexity, the cicada must either transmit more *Hodgkinia* cells to each egg, or risk laying a large proportion of inviable eggs that do not receive all *Hodgkinia* lineages (and thus all *Hodgkinia* genes). Using 16S rRNA amplicon sequencing, we show that *Magicicada* eggs seem to receive all *Hodgkinia* lineages, accomplished by transmitting a larger number of *Hodgkinia* cells to each egg compared when to cicadas with less complex *Hodgkinia*.

We suggest these data show how misaligned selective forces can result in nonadaptive evolution in endosymbioses. In some cicadas, *Hodgkinia* is undergoing repeated lineage splitting in a process that we hypothesize is neutral or selfish from the symbiont perspective and deleterious from the host perspective. A more complex *Hodgkinia* community imposes a cost on the host by forcing it to maintain and transfer larger symbiont masses to each egg. This cost must be borne by the host, since it is completely reliant on its symbionts to survive. In this way, the codependency between host and symbiont facilitates the evolution of nonadaptive symbiont-level traits in endosymbionts when conditions arise where they cannot be prevented by the host.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-296

The inheritance of viable mitochondria

Liliana Milani [1,*], Fabrizio Ghiselli [1]

[1]Department of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy

Abstract: Mitochondria cannot be produced de novo by the cell, but are inherited across generations. Their peculiar genetics (multiple genomes per cell, no meiosis, replication independent from cell cycle, high mutation rate) and the possible exposition to Reactive Oxygen Species (ROS) are predicted to produce a fast accumulation of deleterious mutations, a phenomenon known as Müller's ratchet. Nonetheless, mitochondrial genomes persist accurately over million years. How is a viable mitochondrial genetic information preserved?
To answer this question we review the following relevant topics: 1) the sources of mtDNA mutation (replication and ROS); 2) the origin of mitochondrial membrane potential; 3) the activity of germ line mitochondria; 4) the mitochondrial bottleneck; 5) mtDNA drift and selection. Finally we discuss such topics in the light of an unusual biological system (Doubly Uniparental Inheritance of mitochondria, DUI), in which also sperm mtDNA is regularly transmitted to the progeny.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

OTH-SI6

Symbiosis in the Triatominae subfamily: dynamics, specificity and evolutionary (in)stability of the kissing bug microbiomes

Eva Novakova*, Sonia M R Ruano [1], Vaclav Hypsa [1], Ryan Rego [2]

[1]University Of South Bohemia, Faculty Of Science, [2]Institute of Parasitology, Biological Centre of CAS, Ceske Budejovice, Czech Republic

Abstract: Associations between metazoans and bacteria are widespread phenomena with many important biological consequences, including flexible adaptations to various ecological conditions and life strategies. Using advanced approaches, including confocal microscopy and IIlumina amplicon sequencing we have investigated bacterial associates of the subfamily Triatominae, the vectors of Chagas disease. Although exclusively hematophagous, these important vectors lack obligatory intracellular symbionts. Instead, they host simple microbiomes. Here we bring a complex insight into Triatominae symbiosis characterizing microbiome profiles of US domestic species on the background of several biologically and epidemiologically important aspects. Particularly, using lab colony models, we describe significant changes in diversity driving the microbiome dynamics during the host development. Comparing the lab colonies and field data, we show microbiome specificity attributed to host populations on one hand and apparent evolutionary instability on the other. Regarding the potential impact of microbiomes on vectorial capacity for *Trypanosoma cruzi*, we present our pilot results on *T. protracta* sampled in California, Arizona and Texas. Finally, in order to promote data reproducibility, we identify differences in microbiome profiles sampled with noninvasive and invasive approaches from laboratory colonies.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

OTH-SI3

Does Wolbachia drive the diversity and evolution of sex determination systems in terrestrial isopods?

Thomas Becking, Isabelle Giraud, Maryline Raimond, Bouziane Moumen, Christopher Chandler, Clément Gilbert, Richard Cordaux*

Abstract: In animals, sex differences between males and females are generally determined by genetic factors carried by sex chromosomes. Sex chromosomes are remarkably variable in origin and they can differ even between closely related species, indicating that transitions occur frequently and independently in different groups of organisms. However, the evolutionary causes underlying sex chromosome turnovers are poorly known. In the terrestrial isopod *Armadillidium vulgare*, many lines harbor feminizing *Wolbachia* endosymbionts which can convert genetic males into phenotypic females. Evidence indicates that *Wolbachia* can drive shifts in sex determination mechanisms in *A. vulgare*. In this context, widespread *Wolbachia* occurrence in terrestrial isopods raises the possibility that these endosymbionts may have impacted sex chromosome evolution in terrestrial isopods in general. To test this hypothesis, we characterized sex determination systems in 16 terrestrial isopod species (most of which are infected by *Wolbachia*) using crossings of sex reversed individuals. Next, we established a robust phylogeny of terrestrial isopods using a phylotranscriptomic approach, onto which we mapped sex determination systems. We identified at least four transitions of sex determination systems during the evolution of these taxa. Evidence indicated that YY and WW individuals are viable in many species, suggesting sex chromosomes are at an incipient evolutionary stage. Together, these results support the hypothesis that nucleocytoplasmic conflicts generated by *Wolbachia* triggered recurrent turnovers of sex determination systems in terrestrial isopods. They also establish terrestrial isopods as a model to study evolutionary transitions in sex determination systems.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-277

Correlated accelerations in amino acid substitution rates in the plastid and nuclear genomes of flowering plants

Alissa Williams [1,*], Daniel Sloan [1]

[1]Biology, Colorado State University, Fort Collins, United States

Abstract: Eukaryotic cells contain multiple genomes, and cellular function is often an intricate collaboration between them. For instance, in plants, the core of the plastid CLP protease is composed of nine different types of subunits—one encoded by the plastid genome, and eight encoded by the nuclear genome. Recent work has shown that accelerated amino acid substitution rates in the plastid-encoded CLP subunit correlate with similar accelerations in the nuclear-encoded CLP subunits. However, this previous study only characterized these accelerations within a single genus of plants (*Silene*). In order to determine whether this pattern is more widespread, we extracted the plastid and nuclear CLP genes from a broad range of flowering plants (angiosperms), as well as non-CLP nuclear genes. We then constructed trees for each gene, where branch lengths indicated the rate of amino acid substitution. Using independent sets of the non-CLP nuclear genes to normalize, we compared branch lengths between the plastid and nuclear CLP genes and found a positive correlation. This result indicates that co-accelerations in CLP genes have happened many times independently, highlighting the importance of coevolution between the plastid and nuclear genomes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Synbiosis and interactions

POB-290

A case of evolutionary loss and replacement of an obligate endosymbiont in aphids

Rebecca Chong [1,*], Nancy Moran [1]

[1]Integrative Biology, University of Texas at Austin, Austin, United States

Abstract: Symbiotic interactions between organisms can create new ecological niches and promote diversification. Many insects rely on obligate microbial symbionts to provide essential nutrients limited in their diets. Some symbioses have been established for millions of years, such as that of aphids and their obligate bacterial endosymbiont, *Buchnera aphidicola* (Gammaproteobacteria). Symbionts that transition to intracellular lifestyles often undergo genome degradation, and hosts must adapt and compensate for these deficits. Symbiont genome degradation may compromise function sufficiently to reduce host fitness. In some cases, an evolutionary solution appears to be the acquisition of novel symbionts with intact genomes and functional capabilities. Whereas almost all aphids host *Buchnera*, we found evidence that members of the genus *Geopemphigus* have lost this anciently acquired symbiont and have replaced it with a relatively novel symbiont closely related to *Sulcia muelleri* (Bacteroidetes). We use FISH microscopy to verify presence of the novel symbiont and absence of *Buchnera*. Using Illumina sequencing, we assemble a complete genome of the novel symbiont and reconstruct metabolic pathways to determine if genes that replace functions of the ancient symbiont, such as amino-acid biosynthesis, are maintained. Examples of evolutionary replacement of obligate symbionts provide a valuable opportunity to understand evolutionary and functional genomics of symbiosis.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Systems approaches to behaviour

OM-SA2

Cognitive Ecology of Poison Frogs

Lauren O'Connell [1,*]

[1]Center for Systems Biology, Harvard University, Cambridge, United States

Abstract: Parental care is an important evolutionary innovation, allowing exploitation of novel habitats, influencing fitness and survival of parents and offspring, and serving as an evolutionary precursor to the emergence of social behavior. The elaborate behaviors that encompass parental care are remarkable and require coordinated physiological and neural changes, many of which remain poorly understood. We use poison frogs to understand how evolutionary and ecological forces interact to shape animal behavior. I will first discuss the neural basis of tadpole transport behavior, where we have used comparative approaches among poison frog species that differ in reproductive strategies to gain insight into the mechanisms underlying tadpole piggyback rides. Then, I will discuss how chemical defenses have facilitated the diversification of parental behavior in poison frogs. We have found that the convergent evolution of nursing behavior in South American and Malagasy poison frogs is driven by convergence at the neural and molecular level. This work provides an example of studying the evolution of behavior at various levels, from molecules to organism to ecosystem.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Systems approaches to behaviour

POA-163

The genetic architecture of complex burrowing behavior in Peromyscus mice

Nicole Bedford [1,*], David Zwicker [2], Brant Peterson [1], Michael Brenner [2], Hopi Hoekstra [1]

[1]Organismic and Evolutionary Biology, [2]Applied Mathematics, Harvard University, Cambridge, United States

Abstract: Identifying the genetic changes responsible for variation in complex traits remains challenging, particularly for behavior. Naturalistic behaviors can be difficult to quantify in an unbiased, repeatable, and ecologically relevant way. Here, we focus on heritable differences in behavior between two sister-species of *Peromyscus* mice that build distinct burrow architectures: *P. maniculatus* digs a short simple burrow, whereas *P. polionotus* constructs a complex burrow composed of a long entrance tunnel and upward-sloping escape tunnel. To measure differences in behavior that give rise to these unique burrows, we developed an automated image analysis method that quantifies mouse movement from videos of overnight burrowing in a transparent apparatus in the laboratory. Using this assay, we found that, across trials, most aspects of burrowing behavior (e.g., mouse speed, peak activity time, time spent underground) are highly stable among individuals and distinct between species. Next, by comparing activity patterns between species, we found that the species-specific burrow architectures arise largely from differences in motivation (i.e., time spent digging) rather than ability (i.e., digging rate), suggesting that the causal mechanisms are likely neurobiological rather than morphological. Finally, using a forward-genetics approach, we show that the inheritance patterns and genomic regions associated with mouse activity are often similar to and overlap with those of burrow shape traits, revealing how genetically encoded changes in several mouse behaviors may combine to affect overall burrow structure. Together, we demonstrate how the careful measurement and dissection of complex behavior can shed new light on the genetics and neurobiology of behavioral evolution.

Expanded summary*: Identifying the genetic changes underlying variation in complex polygenic traits remains challenging, particularly for behavior. Unlike many morphological traits, naturalistic behaviors can be difficult to quantify in an unbiased, repeatable, and ecologically-relevant way. Complex behaviors pose a further challenge because they can arise from the integration of multiple behavioral traits, each with distinct genetic underpinnings. Thus, to make connections between genes and behaviors, novel assays are necessary to precisely quantify differences in multiple components of behavior.

Here, we capitalize on extreme differences in burrow shape between two sister-species of *Peromyscus* mice. *P. maniculatus* digs a simple burrow consisting of a single short tunnel, whereas *P. polionotus* digs a complex burrow composed of a long entrance tunnel and upward-sloping escape tunnel. In controlled laboratory tests, mice recapitulate the species-specific burrow architectures of their wild counterparts despite several generations in captivity and no previous exposure to soil, consistent with strong heritability. While the final product constructed by each species is consistent and measurable, the mechanisms underlying the production of such distinct burrow architectures remain unknown. For example, is *P. polionotus* a bigger or stronger mouse, physically capable of longer excavations? Does *P. polionotus* dig more efficiently? Or, is all else equal, and *P. polionotus* simply spends more time digging? These questions are important because their answer can reveal the types of genetic changes that might contribute to variation in burrow architecture. Namely, do the underlying genes primarily contribute to morphological, physiological, or neurobiological differences?

To address this question, we designed a novel video-recorded assay to directly observe the real-time dynamics of burrow construction. We first constructed a narrow, transparent chamber (3'x2'x2") equipped with an infrared illuminator frame to visualize nocturnal, underground behavior. We then developed both an automated video-tracking method to extract several new components of burrowing behavior and a photo analysis protocol that outputs precise measures of final burrow size and shape. Together, these analyses generated 44 traits that provide a detailed and extensive quantification of both mouse behavior and burrow architecture.

Using principal component analysis, we next asked which of these measured traits capture the most variation between *P. maniculatus* and *P. polionotus*. The first principal component (PC1) explained 37% of the total variation and had strong loadings from burrow shape and digging behavior traits such as burrow length, burrow area, time spent digging, and time spent underground. PC2 explained an additional 18% of the variation and had strong loadings from mouse activity traits such as mean speed, time spent moving, and distance covered. Notably, digging rate does not explain a significant portion of the variation between species, nor does mouse morphology. This leaves behavioral—likely motivational—changes as the primary driver of differences in burrow shape.

Next, we took a forward-genetics approach to identify chromosomal regions associated with several components of burrowing behavior. We performed quantitative trait loci (QTL) mapping with 568 F2 males generated from a *P. maniculatus* by *P. polionotus*

intercross. We detected significant QTL for 31 of 44 measured phenotypes. QTL were distributed across 7 chromosomes, and each phenotype had 1-3 QTL peaks. Perhaps unsurprisingly, highly correlated traits (e.g. time spent digging and burrow length) had overlapping QTL peaks. Consistent with previous work, the genetic architecture of burrowing behavior appears relatively simple, with changes in only a few loci contributing to species differences.
Our novel behavioral assay, combined with a genetic approach, allowed us to decompose a complex behavior, both phenotypically and genetically. By doing so, we have discovered new and strong QTL precisely linked to differences in mouse behavior that give rise to differences in burrow architecture. These results lay the foundation for experiments aimed at identifying the particular mutations, pathways, and neural circuits responsible for behavioral evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Systems approaches to behaviour

POA-159

Exclusive pollinators in Acrocomia aculeata (Corozo Palm) from colombian Orinoquia

Astrid Muñoz-Ortiz*, Luis Alberto Nuñez, Lucia Cristina Lozano

Abstract: *Acrocomia aculeata* (Jacq.) Lodd. ex Mart. is a palm known in Colombia like Corozo, its fruits contain high levels of oil used in biodiesel industry and with potential in the production of food and pharmaceutical products. *A. aculeata* is a monoecious and dichogamy palm and its fertilization and fruits production dependent of pollinator insects, conditioning the reproductive success to the presence and the efficiency of pollinators. The objective of this project was to evaluate some biological, ecological and functional aspects of pollinators involved in the fruits production from a field population of Corozo in the colombian Orinoquia. The flower visitors from 16 inflorescences from different individuals of *Acrocomia aculeata* were collected. The pollen contain was evaluated in 320 individuals from 35 species different of insect's visitors. The differences in abundance and frequency, behavior during female and male stages of the visitors did permit to recognize 14 potential species like efficient pollinators. The molecular analysis with the barcode fragment: cytochrome oxidase subunit I, confirmed the taxonomic identification. Additionally, the specificity of pollinators for *A. aculeata* was evaluated using morphological and molecular characters and comparing the presence of the same taxa in others palms species. The results showed that the main pollinators in *A. aculeata* are species from coleopteran families: Curculionidae y Nitidulidae. In particular, the species from genus *Adranthobious* are present in other palms species from genus *Attalea*, in contrast with the species from *Mystrops* genus which were found exclusively in *Acrocomia aculeata*. The results obtain in this project contribute to known, conservation and use of a potential productive system like the Corozo palm. The data associated with the diversity and frequency of pollinators could be the base to improve the industry of this system in the Colombian Orinoquia.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Systems approaches to behaviour

POA-157

Natural loss-of-function gene variants in neuromodulatory pathways among a population of free-ranging rhesus macaques

Michael Montague [1,*], Noah Snyder-Mackler [2], Seth Madlon-Kay [1], Lauren Brent [3], J H Skene [4], Julie Horvath [5], Michael Platt [6]

[1]Department of Neuroscience, University of Pennsylvania, Philadelphia, [2]Department of Evolutionary Anthropology, Duke University, Durham, United States, [3]Centre for Research in Animal Behaviour, University of Exeter, Exeter, United Kingdom, [4]Duke Institute for Brain Sciences, Duke University, [5]Biological & Biomedical Sciences, North Carolina Central University, Durham, [6]University of Pennsylvania, Philadelphia, United States

Abstract: Evidence suggests that variation in social behavior arises from a combination of genetic predispositions and individual experience, yet the underlying biological mechanisms remain poorly understood. To address this gap, we have sought to understand the genetic, developmental, and neurobiological contributions to social behavior in a large, free-ranging population of rhesus macaques (*Macaca mulatta*) with a known pedigree and detailed behavioral phenotypes. We hypothesized that genetic variants underlying molecular differences in neuroreceptors may be associated with behavioral variation in this socially complex species. For example, glutamate receptor interacting protein 1 (GRIP1) is neuronal scaffolding protein involved in stabilization of glutamate receptors at excitatory synapses, and studies of neuronal-specific loss-of-function mice resulted in increased rates of prosocial behavior. To this end, we generated whole genome sequences for 217 individuals and identified over nineteen million population-wide single nucleotide variants. This included 254,777 exonic variants, of which 37% were predicted to alter transcript splicing sites or translated protein sequences. Predicted amino acid changes were found in key genes in neuromodulatory pathways, including dopamine receptors, oxytocin and vasopressin receptors, serotonin transporters, and the mu-1 opioid receptor. Of the 2,022 highest-impact variants, seventeen were predicted to affect candidate genes for autism spectrum disorders, per the online database, SFARI Gene. One such variant, with a population allele frequency of 0.16, was predicted to eliminate the start codon of *GRIP1*. Here, we describe the social behavior among the seventeen heterozygous and eleven homozygous macaques with this variant, suggesting approaches for integrating natural loss-of-function mutations with long-term behavioral data.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Systems approaches to behaviour

POA-160

A transcriptomic study of the Rock Dove (Columba livia) hypothalamic-pituitary-gonadal axis response to stress

Andrew Lang [1,*], Matthew Macmanes [1], Suzanne Austin [2], Rebecca Calisi [2]

[1]Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, [2]Neurobiology, Physiology, and Behavior, University of California, Davis, Davis, United States

Abstract: Stress is a known antagonist of reproductive function. Vertebrate reproduction is generally mediated via the hypothalamic-pituitary-gonadal (HPG) axis, and a large body of work has been dedicated to understanding how stress affects this system. Now, genomic technologies are allowing us to examine the response of the HPG axis to stress at the level of the transcriptome. We investigated the effects of restraint stress on gene expression of the HPG axis using a highly replicated RNAseq study in the model of the rock dove (*Columba livia).* With differential gene expression analyses, we assessed gene expression patterns within and across all tissues, and between the sexes in response to stress. Our results provide the most in-depth characterization to date of the HPG axis in response to stress, offering a genomic foundation for stress and enabling reproductive biologists to better understand reproductive dysfunction in the face of stress.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Systems approaches to behaviour

OM-SA4

Signals of evolutionary history in a learned behavior: song reflects phylogeny in Emberizid sparrows

Nicole Creanza*

Abstract: Approximately half of bird species learn their songs, which generally function in territory defense and mate attraction, and the quality of song learning and performance can affect fitness. Most studies of oscine songbirds focus on the significant song variation within a single species, including spatial variation; this variation raises the possibility that the accumulation of cultural changes over millions of years would obscure signals of evolutionary history. Both birds and humans can often use a bird's song to identify its species; however, it remains unknown whether birds' songs preserve evolutionary information beyond the species level. Here, I introduce computational tools to compare song similarity and genetic relatedness between species; these analyses indicate that features of song evolve at different rates. Analyzing species' song variability in light of genetic distance, I found that certain song features, such as song duration, are good predictors of genetic relatedness, whereas other features are variable within and between species. Mapping these features on genetic phylogenies, I evaluated whether song features evolve at constant rates over time, whether signatures of selection exist in birdsong, and whether correlated evolution has occurred between song features. In addition, I detected evolutionary patterns in individual species' songs over decades of recordings and across geographic distance. In birdsong, there is a tension between the selection pressures for vocal virtuosity and for reliable species recognition: a bird's song should demonstrate to potential mates that he is capable of learning and producing an exceptional song, but he must still be immediately recognizeable as a member of the correct species. These results suggest that these differing selection pressures might act on different aspects of learned song. Combining a genetic framework with a large-scale analysis of learned behavior, these experiments shed light on behavioral variation both within and between species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Systems approaches to behaviour

OM-SA1

Complex Homology and the Neuromolecular Evolution of Social Behavior

Hans A. Hofmann [1,*], Rebecca L. Young [1]

[1]Integrative Biology, The University of Texas at Austin, Austin, United States

Abstract: The social behavior of human and non-human animals can vary tremendously, depending on intrinsic and extrinsic factors, and can be remarkably diverse even among closely related species. Embracing this diversity, behavioral ecologists have provided a fundamental understanding of the adaptive value of many kinds of social behavior and how, and in which ecological contexts, such social systems have evolved. Taking advantage of laboratory animals bred to lack variation, behavioral neuroscientists, in turn, have gained a fairly detailed understanding of how the brain processes and stores socially relevant information, how it generates context-appropriate behavior, and (to a lesser extent) how behavior and its neural substrates develop during ontogeny. Since the beginning of the new millennium, investigators have increasingly become interested in integrating these seemingly disparate disciplines with the goal of (a) unraveling the causes and consequences of individual and population variation in brain and behavior in diverse species; and (b) reconstructing the evolution of the neuromolecular mechanisms that regulate and generate complex behavior. These studies show remarkably conserved roles of hormones (specifically sex steroids and neuropeptides) and neuromodulators (such as biogenic amines) in the regulation of social behavior, even in cases of convergently evolved social systems and across distantly related taxa. Extending these findings on a genomic scale, recent studies provide support for the intriguing hypothesis that coordinated activity of conserved sets of genes underlies independent evolutionary transitions to social behavioral phenotypes. Similarly, neural circuits such as the vertebrate Social Decision-Making Network are highly conserved, suggesting that much of the behavioral diversity we observe in nature reflects variations of an ancient theme. Maybe none of this should come as a surprise: the most recent common ancestor of all animals already had to meet challenges imposed by fluctuating internal states and external environments (finding mates, defending resources, avoiding predators, etc.). The mechanisms used by these ancestral organisms to maintain homeostasis likely served as the building blocks for the evolution of more derived behavioral responses. Here, we will introduce a novel conceptual framework – as well as present results from experimental and comparative studies – that integrate across levels of organization and spatial and temporal scales to untangle the origins and evolution of complex behavioral and neuromolecular phenotypes.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Systems approaches to behaviour

OM-SA5

Complex selection on a regulator of sexual fidelity

Alejandro Berrio [1,*], Rafael F Guerrero [2], Galina V Aglyamova [1], Mariam Okhovat [3], Mikhail Matz [1], Steven Phelps [3]

[1]Biology, University of Texas at Austin, Austin, [2]Biology, Indiana University, Bloomington, [3]University of Texas at Austin, Austin, United States

Abstract: Diversity in social behaviors depends upon standing genetic variation, but we know little about the mechanisms involved. In prairie voles (*Microtus ochrogaster)*, variants at the *Avpr1a* locus predict expression of the vasopressin 1a receptor in the retrosplenial cortex (RSC), a brain region that mediates spatial and contextual memory; cortical V1aR abundance in turn predicts diversity in space-use and sexual fidelity in the field. Here, we examine whether adaptive or neutral forces affect the genetic variation at a putative regulator of V1aR expression. First, we validated balancing selection at the locus by comparing the frequency spectrum of variants at the locus to a random sample of the genome. Next we found strong linkage disequilibrium in the four SNPs that predict high V1aR expression in the RSC, this result suggests that epistatic selection may contribute to maintain their allele associations despite recombination. Analysis of population structure and a haplotype network at two populations revealed that this excessive LD was unlikely to be due to admixture alone. Interestingly, the two populations differed greatly in the regulator of V1aR expression despite the extremely low levels of genome-wide genetic differentiation. Our findings suggest that complex selection on *Avpr1a* locus favors specific combinations of regulatory polymorphisms, maintains the resulting alleles at populations-specific frequencies, and contributes to unique patterns of spatial cognition and sexual fidelity among populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Systems approaches to behaviour

OM-SA6

Systems genomics of bower behavior in Lake Malawi cichlid fishes

Ryan York [1,*], Chinar Patil [2], Kawther Abdilleh [2], Zachary Johnson [2], Matt Conte [3], Tom Kocher [3], Martin Genner [4], Patrick McGrath [2], Hunter Fraser [1], Russell Fernald [1], Todd Streelman [2]

[1]Stanford University, Stanford, [2]Georgia Institute of Technology, Atlanta, [3]University of Maryland, College Park, United States, [4]University of Bristol, Bristol, United Kingdom

Abstract: Unraveling the regulation and evolution of behavior is one of biology's great challenges. Behaviors can be difficult to measure, are ephemeral, and require the coordination of multiple levels of biological organization (i.e. genes, nervous systems, organisms, populations). In addition, very little is known about how these systems vary and evolve in natural systems. Here we use integrated methods to analyze an innate and repeatedly evolved natural behavior: Bower building (mating nest) construction among Lake Malawi cichlid fishes. We find that bower type is associated with extreme genetic divergence via genome-wide association analyses of 20 species possessing dichotomous behavioral traits (pit-type bowers compared to castle-types). Over 800 genes are associated with highly variable genetic variants and are functionally enriched for neural processes such as synaptic plasticity and neuromodulation. RNA-seq analyses in F1 hybrids of a pit-digging species and a castle-building species indicate that over 2,000 genes show evidence of regulatory divergence via allele-specific expression (ASE), that the direction of allelic bias significantly varies with behavioral state, and that there is evidence of lineage-specific selection on the expression of genes associated with synaptic transmission. Furthermore, we find that genes with ASE are significantly more likely to be associated with highly divergent variants, suggesting effects of natural selection acting on the genetic control of expression across bower-building species. Finally, using studies of behavior and immediate early gene expression in the laboratory, we find that bower building may be associated with variation in spatial cognition and that it recruits the activity of diverse brain regions including the dorsolateral telencephalon, the preoptic area of the hypothalamus, and the vagal lobe of the hindbrain. Together these findings suggest that bower building is a highly polygenic trait, is associated with extensive variation in gene regulation, and may rely on intra-species cognitive differences. More broadly our results indicate that even innate and easily measure behaviors may be associated with complex biological bases, supporting the idea that systems-level analyses will be needed in pursuing general principles of behavioral evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Systems approaches to behaviour

OM-SA3

The transcriptomic basis of evolving sexual interactions in Drosophila

Brian Hollis [1,*], Tadeusz Kawecki [2], Laurent Keller [2]

[1]School of Life Sciences, École polytechnique fédérale de Lausanne, [2]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

Abstract: Understanding the evolution of sexual interactions, and in particular the extent to which these interactions represent conflict or synergy between males and females, remains a major challenge in evolutionary biology. In replicate experimental populations of *Drosophila melanogaster*, we have eliminated sexual selection for the last 10 years, corresponding to nearly 200 generations of experimental evolution under randomized monogamy. Males from populations that have evolved without female choice or male-male competition show changes in courtship learning, as well as reduced competitive success, when compared to males from populations that have continued to experience sexual selection. With a series of experiments, we have examined where exactly in the pre- and post-copulatory processes monogamous males fail, and how these males differentially influence female mating and re-mating behavior. We next dissected these phenotypes further by examining evolutionary change at both the sequence level and in the male accessory gland transcriptome and the female head and abdomen transcriptomes, both before and after mating, using a series of crosses between populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Walter Fitch Symposium

OT-WF7

Self-fertility triggers rapid genome shrinkage and loss of sperm competition proteins

Da Yin [1,*], Erich Schwarz [2], Cristel Thomas [1 3], Rebecca Felde [1], Ian Korf [4], Asher Cutter [3], Caitlin Schartner [5], Edward Ralston [5], Barbara Meyer [5], Eric Haag [1]

[1]Biology and Biological Science Graduate Program, University of Maryland, College Park, [2]Molecular Biology and Genetics, Cornell University, Ithaca, United States, [3]Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada, [4]Molecular and Cellular Biology and Genome Center, University of California, Davis, [5]Howard Hughes Medical Institute and Department of Molecular and Cell Biology, University of California, Berkeley, United States

Abstract: Shifts in mating systems can profoundly alter the genome and reproductive traits it encodes. In *Caenorhabditis* nematodes, species that evolved self-fertile hermaphrodites have smaller genomes and reduced mating success than male-female species. The connections between sexual traits and genome size, and the speed with which genome size changes, remain unclear. We compared chromosome-scale genome assemblies for the two most closely related *Caenorhabditis* nematodes with alternative sexual modes, the outcrossing *C. nigoni* and the selfing *C. briggsae*. The *C. nigoni* genome is 19% larger and encodes 24% more proteins, resembling other male-female *Caenorhabditis* and indicating that the *C. briggsae* genome shrank rapidly. The two genomes have similar percentage of repetitive and intronic DNA, while intergenic and coding sequences have become smaller in *C. briggsae. C. nigoni*-specific genes were enriched for those encoding small proteins with male-biased expression. We characterized the *mss* family of sperm surface proteins, found only in outcrossing species and recently lost in *C. briggsae.* Using epitope tagging, we find that MSS proteins enter the secretory pathway and are retained on the sperm surface after activation. In the outcrossing *C. remanei*, *mss*-null males had normal fertility, but their sperm failed to compete with wild-type males. Conversely, restoration of *mss* to *C. briggsae* males was sufficient to render their sperm more competitive than those of wild-type males, and made males more common in mixed-sex populations. These results directly link the reduced mating efficacy of selfing species to the loss of reproductive genes, and highlight the ongoing role that sexual mode plays in shaping the genome.

Expanded summary*: Outcrossing sex is a nearly ubiquitous part of eukaryotic life. However, in animals and plants the costs of sex and scarcity of mates can evolutionarily favor a switch to uniparental reproduction through parthenogenesis or selfing. Such radical changes in sexual reproduction can have equally radical consequences for sexual traits and genome content.

In this research project, we reveal profound consequences induced by the evolution of self-fertility. Our study system is *Caenorhabditis,* the genus of nematodes related to *C. elegans*. Such shifts to uniparental reproduction have evolved repeatedly in both animals and plants, and our results are thus likely more generally applicable across the tree of life. However, *Caenorhabditis* offers an unsurpassed experimental system due to its mature genomic resources and functional genetic tools. We exploit these to make several important discoveries.

We first present a near-chromosome scale *de novo* genome assembly for the male-female *C. nigoni*, and compare it with that for its close relative, the self-fertile *C. briggsae*. These two species are so closely related that they remain partially inter-fertile, yet the *C. nigoni* genome encodes one third more proteins than that of *C. briggsae*. As *C. nigoni* is typical of related outgroup outcrossers, we infer that the genome and proteome of *C. briggsae* has recently and dramatically contracted. More generally, selfing species like *C. briggsae* and *C. elegans* represent shrunken forms of their obligately outcrossing ancestors, with thousands of genes present in the last common ancestor of *Caenorhabditis* now absent in these popular model species. We find that these "lost genes" tend to encode proteins that are shorter and more male-biased in their expression than would be expected by chance, and that much of the loss has come via reduction in gene family size.

Next we identified a large family of genes that encode small proteins that show strongly male-biased and germ cell-biased expression. Within this family, the *male short secreted (mss)* genes form a monophyletic clade found only in outcrossing species and lacking homologs in the selfing *C. elegans*, *C. tropicalis* and *C. briggsae*. In the *C. briggsae* genome, *mss* fragments and a pseudogene indicate recent loss of the *mss* sub-family. We reasoned that such proteins may be involved in reproductive functions that are dispensable after selfing evolves. Using CRISPR-based editing for the first time in an outcrossing nematode, *C. remanei*, we characterize the normal expression and function of *mss* genes. We find that MSS proteins are localized to the surface of activated sperm. A precise deletion of all four *C. remanei mss* paralogs does not prevent normal fertility when all sperm are mutant. However, it causes a profound impairment in the context of wild-type sperm, indicating that MSS proteins are novel sperm competition factors. Remarkably, restoration of MSS expression to males of *C. briggsae* via a transgene is sufficient to confer an advantage over wild-type sperm. Further, MSS+ *C. briggsae* populations maintain higher male frequencies as a result of more effective suppression of selfing by hermaphrodites. This surprising result suggests the provocative hypothesis that the *mss* genes were not simply lost due to relaxed sexual selection, but were actually driven out of the genome by positive Hamiltonian selection to reduce male frequency.

Overall, we believe our study is thought-provoking examination of the ongoing role that sex plays in shaping genome content and organismal traits. It combines genomics, genetics, evolution, and cell biology in a way that should appeal to a broad readership.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Walter Fitch Symposium

OT-WF5

Interpreting Human Genomic Regions Depleted of Archaic Hominin Ancestry

Aaron Wolf [1,*], Joshua Akey [1]

[1]Genome Sciences, University of Washington, Seattle, United States

Abstract: Recent studies have identified archaic sequences in modern human genomes that were inherited from archaic hominin ancestors, such as Neandertals and Denisovans. Strikingly, the distribution of archaic sequence in the modern human genome is heterogeneous, with some large regions depleted of it. Regions that are depleted of archaic sequence may represent loci where archaic sequence was strongly deleterious and rapidly purged from modern human populations. However, alternative mechanisms, such as stochastic loss of archaic sequences due to drift could also contribute to "archaic deserts". To this end, we performed extensive coalescent simulations under a wide variety of demographic models. We find that modern humans are significantly more enriched for large depletions than expected under neutral models. We show that the largest regions depleted of archaic sequence differ from the rest of the genome in several key characteristics, such as being significantly enriched for genes expressed in regions of the brain and differing in their levels of sequence diversity. The largest region depleted of archaic sequence contains the *FOXP2* gene, which is associated with speech and language and carries a regulatory change unique to modern humans. Finally, we leveraged large-scale functional genomics data sets to map putatively deleterious sites Neandertals carried in these regions that may have contributed to the generation of deserts. Understanding the formation and characteristics of regions depleted of archaic introgressed sequence in the modern human genome will help interpret how archaic admixture influenced human evolution and, possibly, what genes may play a role in unique human behaviors.

Expanded summary*: Anatomically modern humans overlapped in time and space with archaic humans like Neandertals and Denisovans. The recent sequencing of the Neandertal and Denisovan genomes has provided insights into human evolution, such as the finding that for individuals of non- African ancestry ~2% of their nuclear genome is Neandertal introgressed sequence and Melanesians carry an additional ~2-4% Denisovan sequence.

We identified introgressed archaic sequence in 503 European, 504 East Asian, and 27 Melanesian individuals using the S* pipeline and the Altai Neandertal and Altai Denisovan reference sequences. Strikingly, the distribution of surviving archaic introgressed sequence across the modern human genome was heterogeneous. While introgressed sequence appeared throughout the genome, several large regions were significantly depleted of it. We also found the overlap of those regions depleted of Neandertal and Denisovan sequence to be significantly greater than expected due to chance.

The size, consistency, and gene content of depleted regions suggest that they arise from common processes. Specifically, we hypothesize that these depletions are products of selection against archaic sequence at these loci. Alternatively, mechanisms such as genetic drift may be responsible for the formation of these features. In previous work, we have examined the probability of depletions of archaic sequence in the modern human genome. We simulated data using well-established models of human demographic history and found that the empirical data have a significantly greater proportion of large (>8Mb) depleted regions than was found in simulated data.

We have since tested further complex demographic models, varying a larger number of parameters in these models. We find that certain parameter sets and model structures are capable of capturing a portion of the empirical distribution. However, no models and no parameter sets examined thus far have been able to fully reproduce the distribution of depleted windows found in the empirical data. The results of these simulations suggest that the empirical data may represent a mixture distribution, and that the formation of these depleted regions is a complex process involving a combination of mechanisms including genetic drift and selection.

We have begun to characterize the largest regions depleted of archaic sequence using large-scale functional genomics data sets. We find that regions depleted of archaic sequence differ from the rest of the human genome in several key measures, such as being significantly enriched for genes expressed in regions of the brain, and differing in their levels of sequence diversity. The largest region depleted of archaic sequence contains the *FOXP2* gene, which is associated with speech and language and carries a regulatory change unique to modern humans. As well, we have identified several loci in these regions that are enriched for fixed differences between human and Neandertal sequence. Some of these loci contain enhancers expressed in neuronal cell lines. We have also identified

several genes that contain non-synonymous fixed differences between Neandertal and human with predicted damaging or deleterious effects.
These depletions of introgressed sequence are unexplored and uncharacterized phenomena that hold important insights into human evolution and the biological differences between modern and archaic humans. Studying these loci, identifying the mechanisms of their formation, and characterizing the features contained within, may be informative in addressing fundamental questions about the evolution of uniquely modern human traits. More generally, our understanding of how modern humans came to flourish in Europe and Asia while other archaic humans perished, remains incomplete. New genetic data and analyses can complement archaeological data, providing estimates of population size, diversity, structure, and migration. Analyzing the genomic remains of archaic-modern human admixture adds to the story of human history, demonstrating the complexity of the interactions between modern and archaic humans.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Walter Fitch Symposium

OT-WF2

Convergent evolution of Y chromosome gene content in flies

Shivani Mahajan [1,*]

[1]Integrative Biology, UC Berkeley, Berkeley, United States

Abstract: Sex chromosomes have formed repeatedly across Diptera from ordinary autosomes, and X chromosomes mostly conserve their ancestral genes. Y chromosomes are characterized by abundant gene loss and an accumulation of repetitive DNA, yet the nature of the gene repertoire of fly Y chromosomes is largely unknown. Here, we trace gene content evolution of Y chromosomes across 15 Diptera species, using a subtraction pipeline that infers Y genes from male and female genome and transcriptome data. Application to *Drosophila melanogaster* data shows that our methodology has high sensitivity and specificity to identify Y genes, and we also discover a novel protein-coding gene on the *D. melanogaster* Y chromosome. The number of inferred Y genes varies substantially between species, and we both identify Y-linked genes in species without morphologically distinguishable sex chromosomes, but also fail to detect Y genes in others with differentiated X and Y sex chromosomes. Young Y chromosomes still show clear evidence of their autosomal origins, but contrary to mammals, most genes on old Y chromosomes in flies are not simply remnants of genes originally present on the proto-sex chromosome that escaped degeneration, but instead were recruited to the Y secondarily from autosomes. Despite no overlap in Y-linked gene content in different species with independently formed sex chromosomes, we find that genes that have been maintained on or recruited to the Y have evolved convergent gene functions associated with testis expression. Thus, male-specific selection appears as a dominant force shaping gene content evolution of Y chromosomes across fly species.

Expanded summary*: Sex chromosomes originated from ordinary autosomes but undergo unusual and unique patterns of genome evolution owing to the lack of recombination on the Y chromosomes. This leads to massive gene loss on Y chromosomes (i.e. Y degeneration), which in turn creates a gene dose imbalance of X-linked genes in males. Even for well-studied model organisms like *Drosophila melanogaster* or humans, the evolutionary forces driving sex chromosome evolution and the molecular processes and mechanisms responsible for their unusual characteristics are poorly understood.

One of the key challenges for studying Y gene content evolution is the highly heterochromatic and repeat-rich nature of the Y chromosome, which makes both its sequencing as well as its assembly extremely difficult. Even for model organisms such as *Drosophila melanogaster* with a tremendous amount of data available, identification of Y-linked genes remains a particularly difficult task computationally. As part of this current study, I have developed a bioinformatics pipeline that combines genomic coverage information with RNA-seq data to circumvent this problem and directly extract putative Y-linked coding sequences without relying on a genomic assembly of the highly repeat-rich Y chromosome . Briefly, I assembled male-specific transcripts from RNA-seq reads that do not align to a female genome assembly and then used a subtraction approach to further narrow down the list of putative Y-linked transcripts. We broadly sampled across the order Diptera and separately sequenced both male and female genomes and transcriptomes from 15 species in order to identify putative Y-linked transcripts in these various species.

This study helps shed light on key aspects of Y chromosome evolution such as: (a) How conserved is the gene content of the Y chromosome across Diptera species, and how much homology is retained with their X-linked counterparts? (b) What global selective forces underlie patterns of gene acquisitions from autosomes to Y chromosomes? (c) What molecular functions do newly discovered putative Y-linked genes have? Moreover, some of the sequenced species have neo-sex chromosomes (i.e. very recently formed Y chromosomes that still display considerable homology with the neo-X), and this study allows a direct comparison between species with neo-Y chromosomes and those that have old, non-homologous Y chromosomes. Additionally for several species we also collected RNA-seq data from somatic tissues as well as testis and ovaries to identify and characterize the differences in patterns of gene expression between somatic tissues and gonads.

The study also has potential biomedical implications because species such as *Anopheles gambiae* and *Aedes aegyptii* that are included in our analyses are known vectors for diseases such as malaria and dengue. Understanding the mechanisms of sex determination in these species, including the genes and pathways involved, is especially important because of the sex-specific transmission of these diseases, i.e. only the female mosquitoes are carriers of malaria and dengue

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Walter Fitch Symposium

OT-WF4

Domestic pigeon's checkered past: wing color pattern variation is associated with one gene, two mechanisms, and interspecific introgression

Anna Vickrey [1,*], Rebecca Bruders [1], Zev Kronenberg [2], Ej Osborne [2], Mark Yandell [2], Michael Shapiro [1]

[1]Biology, [2]Human Genetics, University of Utah, Salt Lake City, United States

Abstract: Birds have evolved a vast array of color patterns in response to natural, sexual, and artificial selection. Rock pigeons (*Columba livia*) vary tremendously in color pattern and are thus a stunning example of this diversity within a species. Four alleles (T-check, checker, bar, and barless in decreasing order of melanism) at a single locus determine the major wing color pattern. Although the bar pattern is likely the ancestral phenotype, checker and T-check birds are more numerous in urban environments, possibly due to enhanced fitness. To investigate the genetic basis of wing color pattern variation, we compared whole-genomes of bar to checker/T-check birds and identified a candidate region that was highly differentiated between the two groups. Surprisingly, sequence comparisons suggest that a haplotype in this region that is shared by all checker and T-check birds was introgressed into the rock pigeon from speckled pigeon (*Columba guinea*), providing a striking example of cross-species transmission of a potentially adaptive phenotype. One gene in the candidate region shows expression differences among bar, checker, and T-check alleles in regenerating feathers, indicating a cis-regulatory change at this locus. Barless birds, which have an increased incidence of vision defects, are homozygous for a coding mutation at this same locus. Remarkably, the same mutation is also observed in two human families with hereditary blindness. In summary, we find unexpected molecular links between color pattern, adaptive introgression, and vision defects, and identify wing pattern phenotypes in pigeons that are associated with both regulatory (checker, T-check) and coding (barless) changes.

Expanded summary*: I am interested in understanding the genetic, developmental, and evolutionary basis of morphological variation. Birds in particular display an enormous amount of morphological variation that has evolved under natural, sexual, and artificial selection. Relatively little is known about the molecular mechanisms that produce this variation. In order to identify genes, specific mutations, and developmental pathways that produce particular traits in birds, our lab uses the model organism rock pigeon (*Columba livia)*. Rock pigeons are an ideal species in which to study morphological variation because of the tremendous phenotypic diversity present among over 350 breeds. Although breeds may look as different from each other as separate species, breeds of rock pigeon are interfertile (allowing for traditional genetics) and their genomes are mostly identical (allowing for tractable genome-wide association studies).

Similar traits in pigeon are present in other species of birds, and once we identify molecular mechanisms that cause a particular trait in rock pigeon, we take a candidate approach to ask whether the same mechanism is responsible for a similar trait in a different bird species. Using this approach, I found that the same gene that causes a feather ornament, called head crest, in rock pigeon is also responsible for causing head crests in a different species; the ringneck dove (Vickrey et al. (2015) *MBE*).

I am currently investigating the molecular basis of pigmentation patterning. Although much is known about the genes and pathways that determine the type of pigment that gets produced (color), very little is known about how the pigments get distributed to produce color patterns. Color patterns are evolutionarily important, and can impact fitness by affecting camouflage, mate-choice, and communication. In rock pigeons, classical genetics determined that four alleles (T-check, checker, bar, and barless in decreasing order of melanism) at a single locus determine the major wing color pattern. Although the bar pattern is likely the ancestral phenotype, checker and T-check birds are more numerous in urban environments, possibly due to enhanced fitness.

We conducted a genome-wide association study and found a single genomic region that is highly differentiated in birds with different patterns. The more melanistic checker and T-check birds in particular had a highly divergent sequence in the candidate region. Pigeon breeders had suggested to us that these patterns might have come from cross-species hybridization with speckled pigeon (*C. guinea*). Surprisingly, we found evidence that a core minimal haplotype shared by all checker and T-check birds may have been introgressed into rock pigeon from *C. guinea*. This provides a striking example of introgression of a potentially adaptive phenotype.

Within the candidate region, we found that additional copies of copy number variable region are associated with more melanistic phenotypes. The CNV lies between two genes and we hypothesized that it might contain a regulatory region. We found that one gene shows expression differences, driven by differential cis-regulation. This same gene is fixed for a non-sense coding mutation in all barless birds, which have an increased incidence of vision defects. The same residue is mutated in two human families with hereditary blindness. Interestingly, wing pattern alleles are associated with both coding and regulatory mutations that act on a common gene.

I am presently working to understand how wing pattern is linked to feral fitness, how and where the candidate gene functions during development, and whether there are other introgressed genomic regions. Our findings inform us about novel gene functions, disease phenotypes common to humans and pigeons, and provide a starting point to identify the basis of similar traits in wild bird species.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Walter Fitch Symposium

OT-WF6

A new theory on the cause of genetic dominance

Xinzhu Wei [1,*], Jianzhi Zhang [1]

[1]Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, United States

Abstract: The cause of the widespread dominance of wild-type alleles over deleterious mutant alleles is a subject of long-standing interest and controversy. Fisher's theory that dominance results from selection is now considered untenable. Wright instead argued that dominance is an intrinsic property of metabolic systems, but his theory cannot satisfactorily explain why dominance is also prevalent in non-enzyme genes. We hypothesize that dominance is a special case of the phenomenon of diminishing returns from advantageous mutations, which means that gaining a wild-type allele at a locus is less beneficial in heterozygous mutants than in homozygous mutants, which is exactly dominance. Our previous work established that diminishing returns results from the modular organization of life where the contribution of each functional module to fitness is determined jointly by the genotype and environment and has an upper limit. To test our hypothesis, we analyze the fitness data of over 7000 genotyped diploids of yeast in nine environments. We find that the fraction (*F*) of beneficial single nucleotide polymorphisms (SNPs) exhibiting dominance exceeds 0.5 in each environment. We define environmental quality (*Q*) as the average fitness of all genotypes in the environment, and find that *F* increases with *Q*. This observation is unexplainable by the existing theories of dominance, but is predicted by the modular life model and is a characteristic of diminishing returns. Furthermore, all previous observations about dominance are consistent with the modular life model. These findings support that dominance is an intrinsic property arising from the modular organization of life.

Expanded summary*: Dominance is among the first phenomena discovered in genetics, yet its cause remains elusive even after a century of investigation. Specifically, it was heatedly debated whether Fisher (Am. Nat. 62: 115–126) or Wright (Am. Nat. 63: 274–279) correctly explained dominance. Fisher believed that dominance results from direct selection for modifiers that increase the dominance and hence fitness of heterozygotes, but this hypothesis is now considered untenable for several reasons, including the extreme weakness of the selection for dominance (e.g., Charlesworth, Nature 278: 848-849). By contrast, Wright argued that dominance is an intrinsic property of metabolic systems. This idea was further developed by Kacser and Burns in their metabolic control theory (Kacser and Burns, Genetics 97: 639-625). While the Wright-Kacser-Burns hypothesis is considered the leading theory on dominance, it cannot explain why dominance is as prevalent in non-enzyme genes as in enzyme genes (Phadnis and Fry, Genetics 171: 385-392) and why dominance remains strong even when assumptions of the metabolic control theory are violated (Marerk and Korona, J. Evol. Biol. 29: 1836–1845). Thus, existing theories cannot satisfactorily explain all patterns of dominance.

A number of experimental evolution studies reported diminishing returns from advantageous mutations, which refers to the phenomenon that the same advantageous mutation is less beneficial when occurring in fitter genotypes. Our unpublished work showed that diminishing returns is also widespread among natural polymorphisms. Our comparison between empirical patterns of diminishing returns and modeling results suggests that diminishing returns originates from the modular organization of life where the contribution of each functional module to fitness is determined jointly by the genotype and environment and has an upper limit.

Here we propose that genetic dominance is a special case of diminishing returns, because diminishing returns means that gaining a wild-type allele at a locus is less beneficial in heterozygous mutants than in homozygous mutants irrespective of the function of the gene involved, which is exactly dominance. We went on to test our hypothesis by analyzing several large datasets of yeast whether dominance can be estimated for many SNPs or gene deletions. We found that the empirical patterns of dominance are similar to previous findings about diminishing returns and thus can be explained by the modular life model. In comparison, previous theories of dominance cannot fully explain these empirical patterns.

We believe this work is important, because it (i) satisfactorily explains dominance, (ii) connects two widespread phenomena in genetics and evolution: dominance and diminishing returns, and (iii) reveals the fundamental importance of the modular structure of life in genotype-phenotype mapping and evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Walter Fitch Symposium

OT-WF1

Does loss of sex lead to increased transposable element accumulation? Evidence from the evening primrose genus Oenothera

David Carlson [1,*], Jesse Hollister [1]

[1]Ecology & Evolution, Stony Brook University, Stony Brook, United States

Abstract: Transposable elements (TEs) are ubiquitous features of eukaryotic organisms that have major impacts on genome architecture and gene regulation. Using a variety of mechanisms to copy themselves, these mobile sequences are able to proliferate across the genome, often at the expense of host fitness. An interesting open question in TE biology is how the loss of sexual recombination impacts TE accumulation. Theoretical predictions and empirical results suggest that the efficacy of selection is reduced in the absence of recombination, potentially leading to increased proliferation of deleterious TEs in asexual species. Alternatively, lack of recombination could produce the opposite pattern, in which TEs evolve lower rates of transposition since the fate of TEs and their host genomes are more closely coupled. The evening primrose genus *Oenothera* is a system well-suited for resolving these competing hypotheses. Due to a genetic system involving chromosomal rearrangements called Permanent Translocation Heterozygosity (PTH), sexual recombination has been independently lost within at least seven *Oenothera* lineages. Using newly sequenced genomic data, we have annotated and quantified levels of TE accumulation in genomes from multiple species, both PTH and non-PTH, sampled from across the *Oenothera* phylogeny. Because we sample from several independent loss-of-recombination events, our dataset represents powerful evidence regarding the influence of recombination on TE evolution.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Walter Fitch Symposium

OT-WF3

Fine-Mapping and Functional Analyses of Genetic Variants Driving Local Adaptations in Humans.

Michal Szpak [1,*], Massimo Mezzavilla [1 2], Qasim Ayub [1], Yuan Chen [1], Yali Xue [1], Chris Tyler-Smith [1] and WTSI Mouse Pipelines, WTSI Mouse Informatics and WTSI Research Support Facility

[1]Human Genetics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom, [2]Experimental Genetics, Sidra Medical and Research Center, Doha, Qatar

Abstract: Identifying the exact variants driving positive selection in the human genome and their functional follow-up is key to understanding the biology of adaptation, and should be the new focus in the area. We have therefore developed a new algorithm for this, Fine-Mapping of Adaptive Variation (*FineMAV*), which combines population differentiation, derived allele frequency and a measure of molecular functionality to prioritise candidate selected variants for functional studies. We calibrated and tested *FineMAV* using eight 'gold standard' examples of experimentally-validated causal variants underlying positive selection, and were able to pick out the known functional allele in all instances. We used this approach to identify the best candidate variants driving local adaptations in the 1000 Genomes Project Phase 3 SNP dataset and report many novel examples, including rs6048066 in *TGM3* associated with curly hair in Africa, as well as rs11150606 and rs201075024 in *PRSS53* linked to hair shape in East and South Asia respectively. We investigated the functions of six genes showing strong signals of selection using mouse knock-outs. The curly whiskers of *Prss53* knock-out mice support the hypothesis of selection in the human ortholog due to hair shape. Finally, we generated nine mouse knock-ins of the human selected alleles. Preliminary phenotyping of the mouse model carrying the selected East Asian allele of *PRSS53* revealed specific head hair abnormalities. Our work is thus facilitating the identification and functional validation of causative alleles driving human adaptations.

Expanded summary*: The genetic basis of human adaptations is of great interest and has a correspondingly large literature. Most previous work has focused on investigating the mode of adaptation and scanning the genome for signatures of positive selection. The current literature thus documents that classic sweeps were not common, and are difficult to identify reliably from population-genetic data alone as attested by the limited overlap between genomic selection scans, but nevertheless have occurred and are of great interest. We have not carried out another genome-wide scan for positive selection and are not entering the debate about whether or not classic selective sweeps were common in humans. Instead, we take the view that the field now needs additional well-supported examples of variants that are driving adaptations, both to understand specific events and to inform more general questions regarding the genetic basis of human adaptation. Support comes most compellingly from model cell/organism studies, but these are low-throughput and so a way to prioritise candidates for them is needed. We provide this by combining population-genetic and functional evidence into a single quantitative measure, the *FineMAV* score, which scans millions of variants genome-wide to generate a list of individual candidate variants in order of priority. We validated our method using a meta-analysis, a handful of gold standard variants, together with available *in silico* evidence for selection. We have begun modelling a few of the candidate variants in mice ourselves, and have reported phenotypic consequences linked to selection in *PRSS53*. We hope that others may also benefit from this work, either directly from the human candidates we identify, or more indirectly by applying our approach to other species, as our method is applicable to any species with suitable genomic data. We thus provide a way to move forward from the morass of genome scans for positive selection. Simulation showed that our method probably misses many genuine selected variants (moderately high false negative rate), but our prioritization aims to enrich for true positives (low false positive rate), which is what matters for people who are going to spend years examining individual candidates in cellular or animal models. *FineMAV* now offers a better way to identify specific variants for modelling and paves the way for identification and understanding of causative alleles driving phenotypic differences among human populations.

Disclosure of Interest: None Declared

Keywords: None

SMBE 2017

Walter Fitch Symposium

OT-WF8

Parallel evolution of influenza across multiple spatiotemporal scales

Katherine Xue [1 2,*], Terry Stevens-Ayers [3], Angela Campbell [4], Janet Englund [5], Steven Pergam [3 5], Michael Boeckh [3 5], Jesse Bloom [1 2]

[1]Genome Sciences, University of Washington, [2]Basic Sciences, [3]Clinical Research, [4]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, [5]Department of Medicine, University of Washington, Seattle, United States

Abstract: Viral variants that arise in the global influenza population begin as *de novo* mutations in single infected hosts. Complex evolutionary dynamics and selective pressures at multiple spatiotemporal scales affect each mutation's fate, but the processes that transform within-host variation to between-host global genetic diversity are poorly understood. Here, we demonstrate that influenza evolution within infected hosts recapitulates many evolutionary dynamics observed at the global scale. We deep-sequence longitudinal viral samples from four immunocompromised patients with long-term H3N2 influenza infections, offering an unusual glimpse into the virus's evolution within these unique human hosts. In the viral surface proteins, a small set of *de novo* mutations arise independently in multiple patients. These mutations emerge repeatedly within single patients and compete with one another in different combinations—a vivid clinical example of the clonal interference frequently observed in experimental evolution. Most of these recurrent within-host mutations also reach high frequency in the global influenza population. Our results demonstrate surprising concordance in evolutionary dynamics across multiple spatiotemporal scales.

Expanded summary*: Viruses rapidly acquire *de novo* mutations as they replicate within infected hosts, but few variants reach a high enough frequency to transmit between hosts and eventually fix on a global scale. Traditionally, influenza evolution has been studied at this global scale, where researchers like Walter Fitch identified extensive positive selection for antigenic evolution. But recent deep-sequencing techniques are making it possible to capture the genetic diversity in the viral population within single infected individuals. Comparing viral evolution at the within- and between-host scales can reveal complex dynamics that are difficult to measure in the lab.

In our study, we sought to better understand influenza's within-host evolutionary dynamics and compare those dynamics to global evolution. We deep-sequenced longitudinal influenza samples from four immunocompromised individuals with long-term infections. These infections provide an unusual window into evolution during what is typically an acute infection.

We found remarkable parallelism in influenza evolution across multiple spatiotemporal scales. Within single patients, we saw that multiple mutations in hemagglutinin, the main target of immune selection, arise independently in distinct genetic backgrounds and compete with one another in different combinations. This clonal interference, which has been observed extensively in experimental evolution, has not previously been reported in a clinical context.

The same mutations also show strong parallelism between patients in our study. In five cases, we observe that the exact same mutations arise independently in two or three patients, which suggests that influenza experiences similar selective pressures even in different hosts, with unusual medical histories.

Lastly, we also find a high degree of overlap in mutations that at the within-host and between-host scales: many of the hemagglutinin mutations that arise within our patients go on to reach high frequency or fix in the global influenza population. Despite heterogeneity in viral and host backgrounds, our results demonstrate surprising concordance in the evolutionary dynamics of influenza across spatiotemporal scales.

Our results have interesting implications for influenza surveillance, in which viral deep-sequencing is becoming more common. We suggest that within-host viral diversity may act as a noisy early measurement of global evolution, shaped by some of the same immunological and evolutionary constraints.

Disclosure of Interest: None Declared

Keywords: None