INTRODUCTION

Studies of DNA polymorphism have been especially intense in the past decades and have yielded ample genomic data for key species and taxa of the evolutionary tree, but have still failed to solve the problem of species and speciation, which is a basic problem of biology. The failure is explained by the discrete nature of a new species as a group evolving genetically independently and the continuous nature of evolution. In the biological concept of speciation in organisms with sexual reproduction is accumulating the differences that suffice to ensure partial or complete reproductive isolation. In turn, the concept suggests unlimited genetic recombination within a species and a discontinuation of gene flow between species [1, 2]. In the case of sympatric populations, which occupy the same area, intercrossing is continuously possible, and even weak gene flow is enough to prevent speciation. Thus, a reproductive barrier must suddenly arise between populations and spread rapidly to allow sympatric speciation.

The objective of this work was to study the relationship between changes in repetitive DNA and the extent of genetic isolation in fish. In addition, we intended to demonstrate in several models that an accumulation and fixation of mutations in noncoding and nonfunctional DNA strongly agrees with phylogenetic reconstructions developed for the groups under study on the basis of morphological traits and coding sequences.

EXPERIMENTAL

Marine, freshwater, and anadromous fish were used as models. Sample collections were kindly provided by the Department of Ichthyology (Moscow State University), Institute of General Genetics (Russian Academy of Sciences), and All-Russia Institute of Fisheries and Oceanography. DNA was isolated from ethanol-fixed tissues (the liver, milt, and fatty fins) according to a published protocol [3] or a Silica method with a NucleoSpin (Macherey-Nagel, Germany) or Diatom™ DNA Prep 100 (IzoGen, Russia) kit. To study the
population genetic structure of the brown trout *Salmo trutta* from rivers of the White Sea basin, scales were used to isolate DNA [4]. The samples and analytical procedures have been described in detail previously [4–8].

To study the divergence of repetitive sequences, we chose the DNA markers that differed in evolution rate, which corresponded to the taxonomic level of the model in question. A microsatellite analysis of the population structure in the Chilean jack mackerel *Trachurus murphyi* was based on published data [9]; genotyping at microsatellites was carried out by PAGE in 8% gel. An analysis of the population structure in the Alaska pollock *Theragra halcogramma* has been described previously [8].

Satellite DNA was examined by a modified restriction enzyme analysis. Fragments obtained by digesting DNA with restriction endonucleases were end-labeled with radioactive isotopes to greatly improve the resolving power of the method (taxonoprint) [10]. The digestion products were resolved by PAGE in 6% gel (20 × 40 cm²) and visualized autoradiographically.

Optimal arbitrary 20-mer primers for multilocus RAPD PCR were selected based on the variation and the reproducibility of amplification and electrophoresis in 1.5–2% agarose gel for each primer.

Because all of our experiments were variants of a fragment analysis, Nei’s genetic distances [11] were used for phylogenetic inferences and were obtained using the TreeConv1.3 program [12] with a bootstrap analysis.

The RAPD PCR products to be sequenced were cloned in *Escherichia coli* with an InsTAclone™ PCR cloning kit (Fermentas, Lithuania). Homologous sequences were search in the NCBI databases using BLAST and BLASTN software [13, 14].

**RESULTS AND DISCUSSION**

**Microsatellite DNA**

Three fish groups of a population level were examined. All of the groups were conventionally considered to be marine (although salmons are known to be anadromous) and to differ in the extent of their isolation. First, a genetic analysis was performed for the Chilean jack mackerel *Trachurus murphyi* from the Pacific. This is the pelagic species that is characterized by long migrations, lack of geographical or other barriers (e.g., the velocity of population movement or hydrographic conditions), and unlimited genetic exchange between populations [15]. Interspecific and intraspecific analyses with only genetic markers have not been described in the available literature [16, 17]. Close species differ in a complex of features (morphological and genetic traits and obligatory parasitic species), but our analysis with three highly polymorphic microsatellite loci (example, Fig. 1a) has not detected signs of genetic subdivision for *Trachurus murphyi* populations (unpublished data).
The Alaska pollock *Theragra halcogramma* was another model group. The species is a commercial subarctic marine fish; is characterized by long distant feeding, wintering, and spawning migrations; and is a shelf species. A minimal extent of geographic and hydrographic isolation is commonly assumed for the species; although the egg transfer is passive, feeding stocks mix and other factors contribute to high gene flow. Microsatellite loci are highly polymorphic in *Theragra halcogramma* (Fig. 1b), but their comparison still reveals certain differentiation of western and eastern Bering Sea populations. Moreover, significant sex dimorphism has been observed on the basis of one of the ten loci for the Sea of Okhotsk population (a collection of 2006), i.e., males had a higher portion of homozygotes, thus deviating from the Hardy–Weinberg proportion. The finding possibly reflects the specifics of demographic processes in the population [8]. Thus, the population structure can be characterized in some cases even when stock isolation is low.

The brown trout *Salmo trutta* from the Kandalaksha Gulf was used as a third model. In this case, a tight association with its habitat (homing), high reproductive isolation, and a low anthropogenic influence (because creeks are poorly accessible) make it possible to study the phenetic diversity and genetic variation of the species both spatially and temporally and to analyze the genetic structure for small populations located only 15 km apart. In addition, microsatellite loci of this anadromous species strikingly differ from highly polymorphic loci of the marine species, improving the reliability of the analysis. Common salmon microsatellite markers are low polymorphic (3–5 alleles) (Fig. 1c) as a result of inbreeding. Still, the populations were in equilibrium and have a certain degree of safety now. Speciation is not observed, but a high extent of genetic subdivision is maintained owing to the salmon living strategy [4].

A more illustrative example of how the structure of microsatellite loci is associated with the extent of duration of population isolation is provided by electrophoretic patterns obtained for locus OMM1070 in two lake isolates of the Far East char *Salvelinus malma krasheninnikovi* from the lakes Glukhoe and Chernoe of the Onekotan Island and the anadromous Kamchatka char *S. malma malma* from the Avacha River. The northern anadromous form and the resident population from the Glukhoe Lake have similarly sized alleles (Fig. 2). Differentiation of the two populations is quite expectable, but a standard allele frequency analysis would be most likely necessary for a statistically reliable estimation of the genetic distance, although the analysis could not be considered perfectly correct because of possible homoplasia.

The other resident isolate (Chernoe Lake) from the Onekotan Island strikingly differs from the other two forms. In 2000, the population was described as a new char species, *S. gritzenkoi*, on the basis of morphological and, especially, osteological traits [18]. While RAPD markers clearly differentiate good species, we performed a RAPD analysis to verify genetically the isolation of the new species. Our results did not confirm that *S. gritzenkoi* is a separate species, but a high level of within-population inbreeding was noted [5]. This preliminary result might provide a weighty argument for the independent taxonomic status of the char form from the Chernoe Lake, needing further investigation.

**Satellite DNA**

A method detecting the restriction sites in DNA [10] was used to study how divergence of extended repetitive DNA sequences is associated with the genetic distances in species and genera of the families Salmonidae and Coregonidae. A study of approximately 50 animal species has shown that band distri-
bution patterns (taxonprints [19]) are species specific, while individual, sex, and between-population polymorphism is not observed. Major bands most likely originate from high copy tandem repeats [11, 15, 20–23].

Concerted evolution via molecular drive is known to result in the high within-family homogeneity of repeats in individuals of a developing group. Many observations confirm that differentiation of local populations due to molecular drive is possible and that, in conditions of at least partial genetic isolation, a repeat family may consequently come to mark the genomes of closely related species [13, 15, 22, 24–27]. We have previously shown that these general concepts are traceable in evolution of repetitive sequences in salmonids of the genera *Salmo*, *Parasalmo*, and *Oncorhynchus* [6]. Homing, which is inherent in true salmon species, results in strong reproductive isolation and subsequent divergence of populations by morphological and molecular characters.

We compared the NJ trees of salmonids differing in evolutionary age and the extent of genetic isolation (Fig. 3). A tree of the genera *Salmo*, *Parasalmo*, and *Oncorhynchus* shows that each genus forms a monophyletic cluster with a good statistical support (Fig. 3a). In particular, this is true for Pacific trout species (the genus *Parasalmo*), which are conventionally isolated in a genus that is in sister relationships to *Oncorhynchus* by Russian researchers and erroneously included in *Oncorhynchus* by the American Fishery Society based on the data [28].

As for the family Coregonidae (Fig. 3c), to which the biological species concept is often thought to be hardly applicable because of free interspecific and even intergeneric crosses (for a discussion, see [29, 30]), a population genetic approach is necessary for studying the genetic structure of one of its genera (the genus *Coregonus*). Yet distinct clades are formed in the tree by the *Coregonus* species to support their morphological separation (Fig. 3c), although the distances are extremely short, as is seen from the branch length in the scaled dendrogram, and internal nodes are poorly resolved and have a low statistical support. American whitefish species is the only monophyletic cluster with a high support (97%). Apart from these species, a high extent of genetic isolation is characteristic of the only species, the Siberian round whitefish *Prosopium cylindraceum*, which is isolated in a separate genus in the family Coregonidae and shows no intercrosses with other species. The separation of the genus *Prosopium* from the other cluster is even more distinct on a UPGMA tree [14].

Salmonids of the genus *Salvelinus* (chars) are a group that is intermediate in the extent of interspecific genetic isolation between true salmons and whitefish species. The group includes many morphological forms and shows intense speciation and form development, so that the taxonomic status of a species has been assigned to many of its resident forms. Considering the lengths of the internal branches in the tree (Fig. 3c), only minor genetic distances separate all questionable species of the *Salvelinus alpinus–Salvelinus malma* complex [31], although geographical isolation is inevitable for the majority of these forms. “Species” are supported statistically only when separated by great genetic distances in the trees. Such species include only the white spotted char *Salvelinus leucomaenis*, which is commonly recognized as a good species, and *Salvelinus svetovidovi*, whose isolation in a separate genus by morphological features [32] is in line with our phylogenetic reconstruction.

The repetitive DNA variation in all of the groups considered agrees not only with the extent of reproductive isolation, but also with the time of clad diversification, which occurred 8–10 million years ago in the case of true salmon species, approximately 16 million years ago in the case of the genus *Salvelinus*, and less than one million years ago in the case of the intermediate *Salvelinus alpinus–Salvelinus malma* species complex. Speciation in the genus *Coregonus* is thought to occur in glacial refugia approximately 15000 years ago [33, 34].

Thus, the internal nodes that support the morphological isolation of species capable of interspecific hybridization are poorly resolved in the evolutionary young group of whitefish species of the genus *Coregonus*, and the genetic distances are low. The findings agree with the idea that the taxonomic ranks are unreasonably high in the group [35, 36]. In view of substantial introgression [37] and the almost total absence of genetic distances [38], the genus *Coregonus* is now often considered to be a network of species [39–41].

As for the genus *Salvelinus* of the family Salmonidae, our findings [5] support the taxonomic status only for *S. leucomaenis* and the endemic species (or genus) *Salvelthus svetovidovi* and agree with the idea that the species status lacks validity in the case of numerous morphological forms of the typical species *Salvelinus alpinus*. The topology of the tree constructed for the family Salmonidae [6], which is characterized by absolute reproductive isolation of its species, testifies that the genera *Parasalmo* and *Oncorhynchus* are wrong to combine in one genus.

**RAPD PCR Markers**

We used a modified PCR protocol [42] with combinations of long (20-mer, rather than 10-mer as in most studies) arbitrary primers to improve the specificity and increase the number of detectable DNA regions. A total of 91 primer combinations were tested, and 15 of them selected for further experiments [5, 7]. Many morphological forms and species of chars of the genus *Salvelinus*, including both valid and questionable ones, were used as a model. Samples were collected from 29 populations of the regions of the Kuril Islands, Sea of Okhotsk coast, Kamchatka, Chukotka,
Fig. 3. NJ trees constructed for salmonids on the basis of Nei’s distances with the TreeConW 3.3b program and presented at the same scale. Bootstrap indices of more than 40 are given. *Salmo salar* was used as an outgroup. (a) The genera *Salmo*, *Parasalmo*, and *Oncorhynchus*. (b) The genus *Salvelinus*. Species of the genera *Oncorhynchus* and *Parasalmo* are shown in the basal part of the tree for a comparison. (c) The family Coregonidae.
The resolving potential of our primer system is characterized in Fig. 4. Theoretically, RAPD may involve any sequences, including both conserved coding and variable repetitive ones. Figure 4 shows the electrophoretic patterns obtained for five char species with several combinations of 20-mer primers. Pooled DNA of all individuals of a group was always used in the reaction. It is clear that polymorphism of the markers is very high and that their relative resolving potential is suitable for characterizing both species and subspecies levels. In parallel with phylogenetic reconstructions, we compared Nei’s absolute genetic distances between commonly recognized good species and questionable ones. Our results demonstrate that Arctic char forms considered to be species by morphological features are not always true biological species. The distances computed using the TreeConw1.3 program were far lower than those between good species.

A tree combining Dolly Varden and Arctic char forms is shown in Fig. 5. Although its resolution is low, the tree has several features of interest. The tree consists of several unsupported blocks, which can be combined by ecological factors (the geographical region and living strategy). For instance, anadromous S. alpinus forms from Finland, Spitsbergen, and Northwestern Russia cluster together. Isolated and endemic forms group in a separate block, and all of the Dolly Varden char forms group similarly. Dolly Varden char forms of the Russian Far East are close to Salvethymus sverovidovi, which is isolated in a separate genus in taxonomy. Its separation is supported by our restriction enzyme analysis of satellite DNA, as discussed above. Internal nodes are supported for the lake isolates of Dolly varden char and anadromous populations from Kamchatka and Paramushir. As expected, the white spotted char Salvelinus leucomaenis forms a well-supported cluster, while the good American species S. fontinalis and S. namaykush group separately to form a basal branch of the tree. The clustering of the two species is not surprising because hybrids resulting from their intercrossing are traceable for up to four generations [43].

A strict conformity between the phylogenetic relationships and geographical localization is similarly seen in chars from five of the Kuril Islands. Separate clusters are similarly formed by different populations of resident isolates or anadromous forms living under similar ecological conditions on one island [7]. Thus, our RAPD marker system is suitable for studying the phylogenetic relationships at the species and intraspecific levels. The fact suggests a variation for the markers, as is characteristic of repeats. Moreover, divergence of noncoding DNA sequences can hardly explain the above tendency to clustering in agreement with ecological conditions.

To study the nature of the markers, both monomorphic and variable fragments were sequenced and tested for homology in the NCBI databases. A BLAST analysis revealed similar sequences mostly in EST databases [13, 14]. A typical search result is shown in Fig. 6 (see full-color insert). Only single RAPD fragments were similar to expressed (i.e., coding) sequences. Homology involved mostly one or two exon regions and a region that lacked similarity to any database sequence (EST, gene, or nucleotide databases), thus belonging to introns or intergenic spacers. The presence of gene regions in the amplification products may reflect traces of adaptive events in RAPD-based phylogenetic reconstructions. Conserved gene sequences cannot ensure a resolution of the polymorphism forms under study, and variable regions are used to construct a tree. Thus, a possible involvement of coding sequences in adaptation does not exclude evolution of repetitive sequences in any way.

Polymorphism of RAPD fragments indicates that the fragments originate from repetitive sequences, but this origin has been demonstrated experimentally for only one fragment as yet. Part of this RAPD fragment, which was amplified from Kuril Dolly Varden char S. malma krasheninnikovi, was similar to cDNA of the Danio rerio epinephrine transmitter protein. Its gene is completely sequenced, and its intron—exon structure is known. One end of our RAPD fragment coincides with the 5’ end of one of the gene exons, while a major part of the fragment lacks homology to the preceding D. rerio intron. We failed to completely sequence the char intron, but DNA reamplification from an exon primer into the intron was performed for several Dolly Varden char populations of the Kuril Islands and showed that the products were the same in size and sequence, with only one exception. One product was almost the same in size as the corresponding RAPD fragment, while the other products were 1.5-fold shorter.

The long and short sequences were compared using BLAST software (Fig. 6). The intron regions were identical in sequence, but with gaps of approximately 50 nt in some regions. Such a structure may suggest a minisatellite, which is a tandem repeat of up to 1000 bp in length and have repeat units of 50–70 bp. Thus, tandem repeats, which seem to occur in introns, show the same properties as pericentric chromatin.

**CONCLUSIONS**

The main objective of this work was to examine fish forms by means of a molecular genetic analysis with noncoding DNA markers, which may reflect the extent of taxon separation at the species and intraspecific levels and at various stages of allopatric or sympatric speciation. The markers (microsatellites, satellites, and DNA regions amplified with arbitrary prim-
Fig. 4. RAPD PCR analysis of char forms and species of the genus *Salvelinus* demonstrates polymorphism of the markers used. (a) Electrophoretic patterns obtained with different primer combinations. (b) Corresponding schemes. A 8-kb DNA ladder and pBR 322/HindIII digest were used as molecular weight markers. Lanes: 1, 2, Dolly Varden char *Salvelinus malma krasheninnikovi* from the (1) South (Iturup and Kunashir) and (2) North (Shumshu, Paramushir, and Onekotan) Kuril Islands; 3, White spotted char *Salvelinus leucomaenis; 4, Salvelinus malma malma* from Kamchatka; and 5, Arctic char *Salvelinus alpinus.* Fragment sizes (bp) are indicated.
Fig. 5. Unrooted UPGMA tree of char forms and species of the genus Salvelinus as constructed on the basis of the RAPD PCR analysis. Bootstrap indexes higher than 40% are shown.

ers (RAPDs)) reflect the specifics of heterochromatin, whose role in reproductive isolation is still unclear and may be important.

While new theoretical models were developed and empirical results obtained to characterize the genetic basis of speciation (for a review, see [44]), it remains unknown what part of the genome is responsible for reproductive isolation [45] and what is a primary factor that triggers rapid genome changes to generate a reproductive barrier. The majority of the models associate postzygotic speciation with a situation where new alleles of coding sequences arise to cause chromosomal incompatibility in meiosis. Although repetitive DNA is no longer thought to be “selfish” or “junk” by many researchers, only a structural and, to a certain extent, regulatory roles are still commonly recognized for heterochromatin [46–48], along with maintaining the information depot and contributing to the organization of transferring genetic information to the next generation.

However, the idea that a reproductive barrier cannot arise without a cessation of gene flow between hybridizing populations is omitted in all of the models. According to Lewontin [49], if the theory of speciation has an element that might be true in all cases, it is the concept that geographical isolation and a strong restriction of gene exchange between populations are a first essential step to speciation. With a low evolutionary rate of the majority of nuclear genes, a mutation may arise and spread through the total population only in conditions of long-term geographical isolation, so that a reproductive barrier cannot form between populations that are in reproductive contact. In the context of Mayr’s biological species concept [1], the only mechanism conceivable for reproductive isolation is based on concerted evolution of extremely simple, extended repetitive sequences. Assuming that tandem repetitive sequences are not subject to selective pressure, they are theoretically capable of mutating and fixing the mutations independently of each other, that is, stochastically. Our study demonstrates that this is not true. All phylogenetic reconstructions based on highly repetitive noncoding DNA follow the consequence of evolutionary events known for the model organisms in question (for a review, see [50]). While the fish groups chosen for the study substantially differ in main evolutionary stages and divergence time [33, 34, 51–55], an integral genome analysis, which simultaneously addresses many anonymous loci, makes it possible to reduce the inequality errors caused by the differences in the molecular evolution rates of macromolecules [56]. Absolute genetic dis-
stances were used as a criterion in the char group, which included almost all of the species and isolated forms (questionable allopatric “species”) [57]. The approach is adequate, as evident from the fact that identical nucleotide sequences were established for RAPD fragments that were amplified with the same primers for different populations and had the same electrophoretic mobility.

Our taxonoprint analysis of Atlantic and Pacific salmon (the genus *Oncorhynchus*) and trout (the genus *Parasalmo*) species showed that the set of DNA fragments was specific for each species or even a form. The results showed additionally that *Parasalmo* is highly separate from the *Oncorhynchus* species [14]. Moreover, significant differences in taxonoprints were observed for sympatric species, which inhabit the same area. The set included Pacific salmons of the genus *Oncorhynchus* (*O. keta, O. gorbuscha, O. nerka, O. tshawytscha, O. masou*, and *O. kisutsch*), two *Salvelinus* species (*S. alpinus* and *S. leucomaenis*), and two chars of the El’gygytgyn Lake (*Salvelinus svetovidovi* and *Salvelinus alpinus*).

All disputable issues of the taxonomy of our model cannot be considered here in more detail because of space limitations. Yet almost all main nodes of the phyllogenetic reconstructions based on taxonoprint and RAPD analyses coincide with those of well-supported consensus trees based on sequences of several genes [58], and allelic diversity of microsatellite and intergenic minisatellite repeats interspaced through the genome corresponds to the extent of genetic isolation of the relevant populations and species. These findings directly implicate repetitive sequences in the establishment of reproductive isolation.

A similar conclusion can be made from the data that X-chromosomal heterochromatin plays a substantial role in establishing the reproductive barrier between two sister *Drosophila* species [59, 60]. Finally, Shapiro and von Sternberg’s [46] concept of genome system architecture suggests that changes in repetitive

---

**Fig. 6.** Graphic representation of the results of searching the NCBI databases for nucleotide sequences similar to RAPD markers. S(a) and S(b) are examples of conserved expressed sequences found in amplification fragments. S(c) is a minisatellite sequence.
DNA play a role in evolutionary diversifications rather than being transmitted through generations without any phenotypic expression. Following these authors, we also think that basic evolutionary events may be initiated in the repetitive genome portion, providing impetus toward the advent of new species.

ACKNOWLEDGMENTS

This work is dedicated to the memory of the prominent Russian biologists B.M. Mednikov, K.A. Savvaitova, and O.F. Gritsenko.

We are grateful to the co-authors of the studies reported previously; researchers of the Department of Ichthyology and All-Russia Institute of Fisheries and Oceanography for collecting unique samples in long-range expeditions; our colleagues for interest in the study, discussions, and help in manuscript preparation for publication; and V.V. Grechko for reading and proof-reading the manuscript and providing helpful comments.

This work was supported by the Russian Scientific Foundation (project no. 14–50–00029).

REFERENCES


*Translated by T. Tkacheva*