# Metadata of the article that will be visualized in OnlineFirst

<table>
<thead>
<tr>
<th>ArticleTitle</th>
<th>Advances of high-throughput sequencing for unraveling biotechnological potential of microalgal-bacterial communities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article CopyRight</td>
<td>The Author(s), under exclusive licence to Springer Nature B.V. (This will be the copyright line in the final PDF)</td>
</tr>
<tr>
<td>Journal Name</td>
<td>Journal of Applied Phycology</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Corresponding Author</th>
<th>FamilyName</th>
<th>Solovchenko</th>
</tr>
</thead>
<tbody>
<tr>
<td>Given Name</td>
<td>Alexei E.</td>
<td></td>
</tr>
<tr>
<td>Division</td>
<td>Department of Bioengineering, Faculty of Biology</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Lomonosov Moscow State University</td>
<td></td>
</tr>
<tr>
<td>Address</td>
<td>Moscow, 119234 GSP-1, Russia</td>
<td></td>
</tr>
<tr>
<td>Phone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fax</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email</td>
<td><a href="mailto:solovchenkoae@my.msu.ru">solovchenkoae@my.msu.ru</a></td>
<td></td>
</tr>
<tr>
<td>URL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORCID</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author</th>
<th>FamilyName</th>
<th>Zaytsev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Given Name</td>
<td>Petr A.</td>
<td></td>
</tr>
<tr>
<td>Division</td>
<td>Department of Bioengineering, Faculty of Biology</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Lomonosov Moscow State University</td>
<td></td>
</tr>
<tr>
<td>Address</td>
<td>Moscow, 119234 GSP-1, Russia</td>
<td></td>
</tr>
<tr>
<td>Phone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fax</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email</td>
<td></td>
<td></td>
</tr>
<tr>
<td>URL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORCID</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author</th>
<th>FamilyName</th>
<th>Rodin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Given Name</td>
<td>Vladimir A.</td>
<td></td>
</tr>
<tr>
<td>Division</td>
<td>Department of Chemistry of Natural Compounds, Faculty of Chemistry</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>Lomonosov Moscow State University</td>
<td></td>
</tr>
<tr>
<td>Address</td>
<td>Moscow, 119234, Russia</td>
<td></td>
</tr>
<tr>
<td>Phone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fax</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email</td>
<td></td>
<td></td>
</tr>
<tr>
<td>URL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORCID</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author</th>
<th>FamilyName</th>
<th>Zaytseva</th>
</tr>
</thead>
<tbody>
<tr>
<td>Given Name</td>
<td>Anna A.</td>
<td></td>
</tr>
<tr>
<td>Division</td>
<td>Department of Bioengineering, Faculty of Biology</td>
<td></td>
</tr>
</tbody>
</table>
Although established biotechnological applications of microalgae e.g., the production of high-value metabolites is based on axenic cultures, exploitation of the mutualistic consortia of microalgae and bacteria quickly comes to foreground, especially in bioremediation and wastewater treatment. This trend shifts the focus from genomic research of certain microalgal species to metagenomic studies of interactions between microalgae and bacteria in natural communities and in artificial consortia. Dissection of the genetic determinants of the robustness and productivity of the consortia become a hot research direction, too. Admirable contribution to this topic had been made by high-throughput sequencing (HTS), while recent breakthrough in this field was entailed by the advent and rapid development of the 3rd generation nanopore sequencing which becomes increasingly accurate while providing unprecedented sequencing performance. Recent progress of the Oxford Nanopore Technologies (ONT) enabled both classical metagenomic analysis of microalgal-bacterial communities based on whole metagenome sequencing as well as taxonomic and genetic profiling based on the amplicon sequencing. The parallel emergence of novel bioinformatic algorithms for processing the metagenomic datasets opened new opportunities for the analysis of structure and physiology of microalgal-bacterial communities. From the practical perspective, the new HTS techniques became a time- and labor-savers in discovery of new microalgae with a high potential for the accumulation of valuable metabolites, biodegradation of hazardous micropollutants, and biosequestration of nutrients from waste streams. Search for prokaryotic species boosting the biotechnological potential of eukaryotic microalgae via mutualistic interactions with them is another important goal. The insights from the both short-read and long-read metagenomics will form a solid foundation for the rational design of microalgal-bacterial consortia for biotechnology. In this review, we briefly outline the benefits of the long-read sequencing for structural and functional investigation of algal-bacterial consortia and summarize recent reports on using this approach for achieving the biotechnology-related goals.

Keywords (separated by ‘-‘) HTS - Nanopore - Amplicon sequencing - Microalgae - Metagenome - Metabarcoding - Profiling - Functional prediction

Footnote Information The online version contains supplementary material available at https://doi.org/10.1007/s10811-024-03267-1.
REVIEW

Advances of high-throughput sequencing for unraveling biotechnological potential of microalgal-bacterial communities

Petr A. Zaytsev¹ · Vladimir A. Rodin² · Anna A. Zaytseva¹ · Maria I. Zvereva² · Alexei E. Solovchenko¹

Received: 31 December 2023 / Revised: 28 April 2024 / Accepted: 1 May 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract
Although established biotechnological applications of microalgae e.g., the production of high-value metabolites is based on axenic cultures, exploitation of the mutualistic consortia of microalgae and bacteria quickly comes to foreground, especially in bioremediation and wastewater treatment. This trend shifts the focus from genomic research of certain microalgal species to metagenomic studies of interactions between microalgae and bacteria in natural communities and in artificial consortia. Dissection of the genetic determinants of the robustness and productivity of the consortia become a hot research direction, too. Admirable contribution to this topic had been made by high-throughput sequencing (HTS), while recent breakthrough in this field was entailed by the advent and rapid development of the 3rd generation nanopore sequencing which becomes increasingly accurate while providing unprecedented sequencing performance. Recent progress of the Oxford Nanopore Technologies (ONT) enabled both classical metagenomic analysis of microalgal-bacterial communities based on whole metagenome sequencing as well as taxonomic and genetic profiling based on the amplicon sequencing. The parallel emergence of novel bioinformatic algorithms for processing the metagenomic datasets opened new opportunities for the analysis of structure and physiology of microalgal-bacterial communities. From the practical perspective, the new HTS techniques became a time- and labor-savers in discovery of new microalgae with a high potential for the accumulation of valuable metabolites, biodegradation of hazardous micropollutants, and biosequestration of nutrients from waste streams. Search for prokaryotic species boosting the biotechnological potential of eukaryotic microalgae via mutualistic interactions with them is another important goal. The insights from the both short-read and long-read metagenomics will form a solid foundation for the rational design of microalgal-bacterial consortia for biotechnology. In this review, we briefly outline the benefits of the long-read sequencing for structural and functional investigation of algal-bacterial consortia and summarize recent reports on using this approach for achieving the biotechnology-related goals.

Keywords HTS · Nanopore · Amplicon sequencing · Microalgae · Metagenome · Metabarcoding · Profiling · Functional prediction

Introduction: Microalgal consortia as a promising vehicle for biotechnology

In nature, microalgae exist within microbial communities with other microbial species including diverse fungi, bacteria, and/or archaea. In these communities, microalgae become engaged in a complex network of interactions with their partner species represented mostly by bacteria, implemented as trophic exchange and/or chemical signaling. There is ever increasing evidence of the correlation between composition and activity of the bacterial component of the consortium and the physiological condition of the microalgae. This evidence suggests that the interactions between the microalgae and the bacteria can be significant for the
There are species and whole taxa of microalgae whose microbiomes are of a considerable interest due to their high biotechnological potential or even a possible threat to human health and economics (Kublanovskaya et al. 2020b; Danish-Daniel et al. 2023). Among the most conspicuous forms of microalgal-bacterial interactions, and hence most studied so far, is the formation of complex structures such as flocules and biofilms or their biomimetic analogs—photogranules (Trebuch et al. 2020, 2023). They frequently include prokaryotic oxygenic phototrophs—cyanobacteria (Kublanovskaya et al. 2019, 2020a).

A crucial role in the formation and evolution of microalgal-bacterial consortia is played by the phycosphere. This term was coined to denote a spatial zone in close proximity to the microalgal cell surface characterized by the presence of superficial structures of microalgal cells as well as by gradients of chemical and physical parameters making the phycosphere especially favorable for other organisms. In other words, microalgae acts as ecosystem engineers or, in terms of ecology, edificator of the microbial community formed around its cells. Eventually, the phycosphere becomes inhabited by microorganisms engaged in diverse (mostly symbiotic) interactions with the basibiont (the microalga) and between themselves (Fig. 1).

These interactions can be significant from the practical standpoint (Seymour et al. 2017). The most known is successful application of mixed cultures of microalgae with plant growth-promoting bacteria (PGPB) for soil remediation and biofertilization (Gonzalez and Bashan 2000; de-Bashan et al. 2021; Gonzalez-Gonzalez and de-Bashan 2023). The co-culture improves soil health and stimulates crop plants productivity by synthesizing a broad spectrum of bioactive molecules (de-Bashan et al. 2004; 2021) including the phytohormone analogs excreted by representatives of Chlorella, Scenedesmus, and Chlamydomonas.
Cyanobacteria also fix nitrogen (Llamas et al. 2023) and make it, together with phosphorus, more bioavailable to crop plants with participation of microorganisms from the genera *Azospirillum*, *Azotobacter* and other diazotrophic cyanobacteria (Scognamiglio et al. 2021; Solomon et al. 2023). Consortia of microalgae and plant growth-promoting bacteria (PGPB) boost growth and pathogen resistance of important vegetable crops including tomato, onion, and cucumber by stimulating their nitrogen uptake and producing bioactive polysaccharides (Kang et al. 2021). With cyanobacteria added to a microalgae-PGBP consortium, a robust synthetic consortium is formed which can serve as efficient biofertilizer (Sadvakasova et al. 2023). A similar result could be achieved by co-immobilization of microalgae-PGBP consortia on alginate and/or chitosan beads (Gonzalez and Bashan 2000).

Another major application field for microalgal-bacterial consortia is the biotreatment of wastewater by biosequestration of nutrients, decomposition of bulk pollutants, and biodegradation of hazardous micropollutants. Common issues of the microalgal-based solutions for environmental applications including their stability and sustained efficiency under fluctuating environmental conditions and wastewater composition, as well as economic viability can be, in principle, addressed by appropriate microalgal-bacteria consortia (Saravanan et al. 2021).

Bacteria from certain taxa, frequently belonging to PGBP as well, also exert stimulatory effects on microalgal growth and productivity. In analogy with PGBP, those bacteria were named microalgal growth-promoting bacteria (MGBP). Supplementation of MGBP to axenic cultures of microalgae from the genera *Chlorella*, *Chlamydomonas*, and *Euglena* frequently used in wastewater treatment increase biomass accumulation and the treatment efficiency (Toyama et al. 2018).

The most robust form of the algal-bacterial consortia in the wastewater treatment systems are algal-bacteria biofilms (Clagnan et al. 2023). These biofilms can be formed with participation of quorum-sensing mechanisms orchestrating the microalgal-bacterial interactions to attract the MGBP to populate the niches formed around the photoautotrophic cells (Qixin et al. 2022). The MGBP can either stimulate the growth of microalga by supplying them with essential co-factors and vitamins (Shetty et al. 2019; Iqbal et al. 2022) or perform enzymatic hydrolysis of the microalgal cell wall increasing the product yield in case of valuable metabolite production (Carrillo-Reyes et al. 2016). Increasing the bioavailability of nitrogen by bacteria in wastewater sludge communities facilitates accumulation of microalgal biomass (Leong et al. 2020) and, in certain cases, lipid productivity of species from the genera *Chlorella*, *Chlorococcum*, *Scenedesmus*, and *Nannochloropsis* (Koreivienė et al. 2014; Arutelsvan et al. 2021; Upadhyay et al. 2021). Future breakthroughs in wastewater treatment are expected from application of multi-omics approach and high-throughput methods for screening for selection and/or design of even more robust and productive consortia (Patel et al. 2017; Pdamaperuma et al. 2018; Nagarajan et al. 2022).

Clearly, the environmental and agricultural applications of microalgae are about “xenic” cultures and consortia. Moreover, the advent of molecular methods of culture purity control revealed that many microalgal cultures that passed conventional axenicity tests appeared to be not really axenic and harbored other (non-cultivable) microorganisms. Interestingly, using strictly axenic cultures in microalgal biotechnology was frequently complicated by deterioration of culture vigor and productivity, let alone the costs of axenicity maintenance at large scale (Patel et al. 2017; Pdamaperuma et al. 2018).

These circumstances have focused interest to the consortia themselves and methods of their investigation and engineering. It became clear that engineering of the phycosphere aimed at to populating it with desirable MGBP would ensure a kind of division of labor between the components of the consortium for avoiding metabolic overload, enhanced biomass accumulation, balancing the growth by quorum sensing mechanisms, and increase of nutrient availability for microalgae (Park et al. 2017; Patel et al. 2017; González-González and de-Bashan 2021). Specific examples include significant increase of chlorophyll, lipid, and carotenoid content in co-cultures of microalga with the bacteria that are frequently found in microalgal core microbiome such as *Paracoccus haenndaensis* – *Lactobacillus fermentum*, *Characium* sp. – *Pseudomonas composti*, *Tetradesmus obliquus* and *Coe lastrella* sp. – *Variovorax paradoxus* (Berthold et al. 2019; Choi et al. 2021; Perera et al. 2021).

Of special interest is boosting the productivity of the biotechnologically important microalgae such as *Haematococcus lacustris* without resorting to their genome modification. Solving this problem would make the natural astaxanthin from microalgae much more competitive than it is now. Thus, *H. lacustris* has shown more than two-fold increase in its major secondary carotenoid astaxanthin yield in co-culture with the bacteria *Sphingomonas hankookensis* or *Paenarthrobacter ureafaciens*, or the fungus *Simplicillium lansononiveum* (Lee et al. 2022). Co-culturing of microalgae with certain yeast species results in beneficial cross-feeding that either increases the rate of carbon dioxide assimilation or enables the utilization of organic carbon sources for higher biomass accumulation (Cheirsilp et al. 2012; Wang et al. 2016; Gao et al. 2023b). Such co-cultures are designed by high-throughput screening of suitable auxotrophs among microalga, bacteria, and fungi to arrange the most efficient trophic interactions (Saleski et al. 2019).

Taking a closer look on the publication landscape related with microalgal genomics, one might notice that...
molecular biology methods have become widespread within that field (Fig. 2, see also online Supplementary). However, the topic of high-throughput sequencing (HTS) is still underrepresented for microalgal biotechnology. But even then, the initial focus of sequencing techniques on microalgal metabolism and culturing has been gradually shifting towards the genomics approach of microalgal-bacterial community investigation. As a result, HTS methods are becoming tightly related to environmental research, such as ecological monitoring of phytoplankton (including notorious algal blooms and eutrophication), aquaculture, and wastewater treatment. Today we see the emerging understanding of the importance of HTS for monitoring of microbial and microalgal diversity, as well as for estimation of microalga-bacterial consortia functional potential. At the same time, it became clear that solving these problems demands new experimental methods and data processing algorithms in metagenomics. Below, we attempt to outline the importance of long-read sequencing for getting insights into the structure, functioning, and biotechnological potential of algal-bacterial consortia. 

Pro et contra of the mainstream sequencing technologies will be discussed with an emphasis of nanopore sequencing represented by Oxford Nanopore Technology (ONT). Special attention will be given to novel algorithms developed for gaining actionable insights from the data output of ONT sequencing platform. The review covers the reports (Fig. 2, see also online Supplementary) on the successful applications of HTS in the field of microalgal ecology and microalgal-bacterial interactions in the context of biotechnology. Additionally, the amount and specificity of the long read-based metagenomics is considered.

Fig. 2 Co-occurrence map of the keywords in publications related to microalga genomics. The largest number of edges are within ‘molecular methods’ cluster (brown) and to ‘symbiosis’ node of ‘metagenomics’ cluster (yellow), stating the increasing attention to microalga-bacteria interactions in molecular systems biology field. At the same time, the nodes ‘metagenomics’ and ‘microbiome’ have few and thin edges with ‘bioproduction’ and ‘microalga technology’ clusters (purple and blue, consequently), which highlights future potential of metagenomic studies for practical application of alga. The initial set of titles, abstracts, and keywords of 767 research articles was collected from PubMed and analyzed in VOSviewer 1.6.20 (only the keywords which occurred 10 times and more were taken, clustering resolution = 1.3, min. strength = 8).
Metagenomics in microalgal research

Metagenomics is a key to knowledge of the microbial universe

Metagenomic approach to investigation of microbial communities evolved in the last two decades. It became a powerful tool for studies of the microbiomes of soil, marine and freshwater sediments, and planktonic communities, as well as microorganisms of animals and plants. This approach also proliferated into diverse practical applications such as environmental monitoring, control of food quality and fermentation, medical research, and wastewater treatment. Recently, researchers started to use metagenomics to dissect microalgal-based communities from various biotopes from active sludge of wastewater treatment plants to photobioreactors. An illustrous example is comprised by culture crash “forensics” (Lane et al. 2013).

Metagenomic approach becomes increasingly widespread in the studies of microbial communities while the classical methods that are based on isolation and cultivation are giving up their positions since the latter are (i) labour- and time-intensive and (ii) suffer from a high organism-dependent bias. An important advantage of metagenomics is its potential to reveal hidden microbial diversity represented by uncultured species. This is especially relevant to bacterial symbionts of microalgae in natural and artificial systems.

Generally, metagenomic studies of alga-bacterial communities aim to answer three practical questions:

1. What organisms form the community (which taxa do they belong to)?

This question is solved using molecular identifiers or barcodes uniquely identifying organisms at different levels of taxonomy. According to the principle of DNA barcoding, sets of genomic loci are selected to ensure the desired level of identification accuracy for bacterial and microalgal strains. While the 16S rRNA gene locus is usually sufficient for identification of the most of heterotrophic bacterial species in the microalgal phycosphere (Lebonah et al. 2014), reliable identification of oxygenic phototrophs requires a more extended set of loci. Thus, for eukaryotic microalgal nuclear genes (18S rRNA, nuITS1, and nuITS2), chloroplast genes (rbcL, tufA, and cp23S), as well as mitochondrial cytochrome c oxidase subunit I (COI) gene are used in most situations (Hadi et al. 2016; Zou et al. 2016; Ballesteros et al. 2021). Among those, the tufA gene encoding a plastidial elongation factor currently is the most promising marker capable of resolving lower taxa within the class Chlorophyceae (Vieira et al. 2016).

For identification of Cyanophyta, the 16S rRNA gene and ITS between 16S and 23S rRNA genes, functional rbcL or nif genes, and a subunit of RNA polymerase (rpoB/CID genes) are commonly used (Mishra 2020; Ballesteros et al. 2021). The CBOL (Consortium for the Barcoding of Life) recommends the consequent application of at least two markers for reliable identification of microalgal taxa (Pawlowski et al. 2012).

2. What is the potential functional profile (ecological function) of the community?

The possible physiological and other features of a community are defined by list of functional orthologs represented in the genomes of species forming this community. A more or less specific set of genetic determinants can be compiled for any major phenotypical trait expressed at the level of community. Typical examples include (but not limited to) nitrification (amo, nxr, hao, etc.), denitrification (nap, nar, and nirS, etc.), and uptake of phosphate with its subsequent accumulation in form of polyphosphate (pho genes, PSR1, PTC1, ppk, ppk2, etc.) (Wang et al. 2023; Xiong et al. 2023). Following the concept of reverse ecology, such gene sets might be the basis of metabolic reconstruction of an entire microalga-bacterial community (Cao et al. 2016). This approach might also reveal a lot of information about the biotechnological potential of a community which might be useful e.g. for in silico pre-screening.

3. What are the possible interactions between the organisms forming the community?

Answering this question requires study of the genes responsible for different modes of communication between microalga and bacteria in the phycosphere, from trophic interactions to chemical signaling based on specific molecules (see e.g., Fig. 1). Excellent examples of the latter are quorum sensing substances, phytohormones, algicides, growth inhibitors, and extracellular enzymes that modulate their activity in the medium (Dow 2021; Astafyeva et al. 2022; Santo et al. 2022). Another route of interactions within the community is horizontal gene transfer between species—the phenomenon noticed in natural microalgal-microbial communities under selective pressure of hazardous micropollutants (Liu et al. 2022; Li et al. 2023a).

HTS in microbial community research: pro et contra

Since the advent of the first method for sequencing nucleic acids, this approach has evolved dramatically yielding three generations of sequencing with distinct advantages and drawbacks (Table I; Sanger and Coulson 1975; Slatko et al. 2018). Each method has its own unique characteristics and...
Table 1 Three generations of nucleic acid sequencing methods generally applied in microalgal-bacterial community research

<table>
<thead>
<tr>
<th>Generation</th>
<th>Sequencing method</th>
<th>Platform</th>
<th>Underlying principle</th>
<th>Advantages</th>
<th>Drawbacks</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>First generation sequencing</td>
<td>Chain termination method (Sanger sequencing)</td>
<td>– Agilent Bioanalyzer</td>
<td>Synthesis of a new DNA chain, a complementary matrix chain, with the inclusion of labeled nucleotides stopping the synthesis process</td>
<td>High accuracy</td>
<td>Low throughput, high costs, inefficient for sequencing large genomes and metagenomes</td>
<td>Suitable for: sequencing of individual genes (short fragments), verification of the results of other methods, identification of axenic cultures</td>
</tr>
<tr>
<td>Second/Next generation sequencing (SGS/NGS), High-throughput sequencing* (HTS), Sequencing by synthesis</td>
<td>Pyrosequencing – Roche/454</td>
<td>Measurement of the released pyrophosphate during the incorporation of nucleotides into a growing DNA chain</td>
<td>Longer reads</td>
<td>Errors in sequencing homopolymer regions, higher error in bases, sensitive to DNA quality</td>
<td>Good for amplicon sequencing, suitable for detecting variations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dye sequencing – Illumina</td>
<td>Sequencing by synthesis (incorporation of designated nucleotides and fluorescence detection)</td>
<td>High throughput, low cost of reading, high accuracy</td>
<td>Shorter reads (up to 250 b.p., difficulties in resolving long repeats</td>
<td></td>
<td>A mainstream platform for genomics and transcriptomics of axenic cultures</td>
</tr>
<tr>
<td></td>
<td>Semiconductor sequencing – Ion Torrent</td>
<td>Sequencing by synthesis (detection of hydrogen ions released during incorporation of deoxyribonucleotide triphosphates)</td>
<td>High throughput, low cost of reading</td>
<td>Reads up to 400 b.p. (errors in sequencing homopolymers)</td>
<td>Suitable for microbial genome and transcriptome sequencing, targeted sequencing</td>
<td></td>
</tr>
<tr>
<td>Third generation sequencing (TGS),</td>
<td>Single-molecule real-time sequencing (SMRT) – PacBio</td>
<td>Sequencing by replication with labeled nucleotides</td>
<td>Long reads</td>
<td>Low throughput, relatively high costs and relatively scarce availability worldwide</td>
<td>Fits well for closing gaps in reference assemblies and characterization of structural variation in genomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nanopore sequencing – Oxford Nanopore Technolog (ONT)</td>
<td>Recording of conductivity changes during movement of nucleic acid molecule through a nanopore</td>
<td>Long reads (tens thousands b.p.), portability, high throughput, low cost of reading</td>
<td>Relatively low accuracy</td>
<td>A promising platform for metagenomics and transcriptomics of xenic cultures and communities from environmental applications</td>
<td></td>
</tr>
</tbody>
</table>

*Depending on a particular ONT product the nanopore sequencing can also be considered as HTS
the method of choice depends on specific goals and requirements of the study. Instead of reviewing technical details of each method (for those, we refer the reader to recent overviews: (Mardis 2017; Slatko et al. 2018)), here we will highlight their scope and applicability for investigation of microalgal-bacterial communities with emphasis on the most recent long-read-based technologies.

The third-generation sequencing is distinguished by its ability to read long DNA sequences. This method is represented on the market by single-molecule real-time sequencing technology (SMRT) from PacBio and by solutions from Oxford Nanopore Technology (ONT) company. Admittedly, PacBio has proven itself as the most powerful sequencing method existing up to now, due to the optimal ratio of potential read length to sequencing precision. Still, it has not become a mainstream technology, particularly due to its high cost (almost twice as much compared to other TGS) that leads to lower availability (Athanasopoulou et al. 2022).

However, the second technology is becoming more widespread due to the opposite trends. In the process of nanopore sequencing, individual single-stranded DNA or RNA molecules pass through the nanopore causing changes in electrical conductivity in the nanopore unit. The corresponding changes in the electrical signal are recorded, and the nucleotide sequence is inferred from these records. Further details on the principles and technical implementation of this method can be found in (Kasianowicz et al. 1996; Stoddart et al. 2009).

A key advantage of TGS is the length of reading, which can reach 100 thousand b.p. which is unattainable for other sequencing platforms. Another decisive advantage of TGS is its ability to perform single-molecule sequencing without the need to average signals from a group of molecules making the results more accurate. Combination of these advantages makes the sequencers capable of precise analyzing of long repeats and GC-rich regions of the genome, unlike the alternative technologies (Jain et al. 2018). Furthermore, nanopore sequencing, provided that a sufficient amount of genomic DNA (above 200 fmol) was extracted from the biological sample, makes it possible to omit the amplification step in the sample preparation routine, while PacBio requires pre-amplification for some purposes (Athanasopoulou et al. 2022). The lack of the PCR amplification step lowers the risk of error and sequencing bias. In the case of ONT, it also increases the throughput of reads by reducing the risk of PCR bias and lowering the risk of errors in the detection of biological reads.

At the same time, this technique from the very beginning suffered from a larger error rate. For example, the 10X Genomics Chromium platform was proved (Khrenova et al. 2022). As a result, nanopore sequencing now has a broad range of applications including genomics, epigenomics, metagenomics, and RNA research. It is widely used in life sciences research, medicine, agriculture, and other fields where nucleotide sequencing or nucleotide sequence analysis is required (Zhang et al. 2022; Badger et al. 2023; Mastrorosa et al. 2023).

One of the distinguishing features of ONT is the direct nucleic acids sequencing ability, which opens new opportunities for high accuracy transcriptomics, including identification of novel isoforms and detection of full-length RNA (Athanasopoulou et al. 2022). On the other hand, the accurate differential analysis based on long-read sequencing may require higher throughputs via generation of cDNA library, which in case of nanopore sequencing still provides an advantage over existing methods by reading full-length isoforms and avoiding (or at least reducing) additional bioinformatics step to assemble reads into transcripts. However, this potential has not been fully leveraged due to the limitations of current long-read assembly methods and underdeveloped short-read data integration approaches. Unusually low coverage when using short-read technologies leads to the splitting of one transcript into several transcripts or incorrect definition of ends and, as a result, to errors in the assessment of differential gene expression. Conversely, long-read sequencing libraries lack depth of coverage and suffer from artifacts in cDNA-based methods, leading to erroneous assembly and quantification of transcripts. To overcome these problems, a hybrid assembly approach (short and long reads together) is used, which dramatically increases the sensitivity and accuracy of full-length transcript assembly on the correct strand and improves the detection of biological features of the transcriptome (Kainth et al. 2023). When alternative splicing has a significant contribution to transcriptomic variation, ONT protocols have been shown to be superior to short-read sequencing protocols in terms of transcriptome assembly and the risk of false positives due to unambiguous mapping of reads to transcripts (Engelhard et al. 2023).

Both PacBio and ONT are suitable for implementation of two main strategies of metagenomic studies: whole...
metagenome sequencing and amplicon sequencing of a specific loci either for identification of the microbes and/or revealing their functional potential (Athanassopoulou et al. 2022; Kim et al. 2022). Thus, in 16S-based studies, PacBio and ONT allow the creation of primers covering the entire 16S rRNA gene or even entire ribosomal operons, increasing dramatically the resolution of the taxonomical assignment i.e., the number of precisely distinguishable species (Kerkhof et al. 2017; Tedersoo et al. 2018). Reading the whole metagenome leads to minimal bias in species composition and amount. At the same time, amplicon sequencing of DNA-barcodes (or metabarcoding), e.g., 16S rRNA, its internal transcribed spacer (ITS), rbcL etc., offers a cheaper alternative which features a higher throughput but is potentially prone to bias due to the presence of amplification step (see above).

It is well known that “traditional” short-read sequencing technologies cannot reliably resolve repeats and duplicated regions of the genome, so their using for taxonomical assignment and genome assembling of closely related species is complicated (Ashton et al. 2015), while heterogeneity inherent in the metagenome might lead to incorrect assembly between species. In case of metabarcoding, the 16S rRNA gene sequence harboring a combination of conservative and highly variable regions allows for precise species identification, but limitations of the short-read technologies (NGS, Table 1) prevent them from covering a sufficiently long part of this gene to provide species-level resolution (Shin et al. 2016).

Nowadays, TGS (mostly nanopore sequencing) has secured its place in the array of methods for studies of microbial communities offering distinct advantages for metagenomics. Despite some admirable results produced by PacBio technology in assembling whole genomes of microorganisms, including microalgae (Luo et al. 2018; Maeda et al. 2019; Gao et al. 2023), there are few works dedicated to PacBio evaluation of microbial communities (Tedersoo et al. 2018; Gueidan et al. 2019; Kim et al. 2022). Therefore, we shall consider below the specific applications of the nanopore technology (solely or in combination with short-read methods) for scrutinizing the microalgal community structure and functional profile.

**Studying the whole metagenome of microalgal communities with HTS technologies**

The whole metagenome sequencing (WMS) approach stands as the golden standard for metagenomic studies of various sample types harboring microalgal-bacterial consortia, mostly due to the large amount of sequence data enabling thorough analysis of the consortia. That includes precise taxonomical identification of eukaryotic and prokaryotic species forming a community, confirming the presence of diverse functional genes sets, search for new efficient and stable enzymes and reconstruction of metagenome-scale metabolic models (Belcour et al. 2020; Zorrilla et al. 2021; Kuppa Baskaran et al. 2023). Further insights can be obtained by investigating raw metagenome reads or scaffolds, for example from phylotyping based on straightforward count in alignment-free algorithms (Inskeep et al. 2013; Patil and McHardy 2013), more precise taxonomical identification by BLAST or another sequence comparison tool such as implemented in MEGAN or TAXAssign algorithm (Huson et al. 2007; Inskeep et al. 2013), or classification based on the species-level genome bins e.g., with MetaPhlan 4 algorithm (Ljaz and Quince 2013; Blanco-Míguez et al. 2023).

The most popular approach relies on pre-assembled genomes from the metagenome (MAG) for prokaryotic and eukaryotic species, which however might be limited by insufficient coverage of taxa and quality of the assemblies (Yang et al. 2021). Though application of both mentioned approaches is better adopted for prokaryotic species, there is an emerging trend in algorithm development for eukaryotic microorganisms, including microalgae. Such tools as EukRep and Tiara utilize machine learning and deep learning methods to classify read subsets that are related to a microalga (or even its plastids and mitochondria) in a whole metagenome, then extract and assemble them (West et al. 2018; Karlicki et al. 2021). Completeness and contamination are two main characteristics of MAGs, which are estimated by single-copy marker gene analysis (SCMG). For prokaryotic MAGs, the CheckM algorithm is widely used and shows good performance, while quality check of eukaryotic MAGs is a challenge, it is however reached by using a defined set of eukaryotic SCMG (BUSCO and CEGMA algorithms) or dynamic selection of an appropriate SCMG set for improved evaluation e.g., with EukCC algorithm (Saary et al. 2020).

Though the short-read WMS inherently provides excessive metagenome coverage, its results are still limited by the read length. Confident assignment of the metagenomic reads to a specific taxon by comparison with known DNA barcodes or reference genomes requires longer sequences than obtainable with currently available NGS platforms (Table 1). The robustness of genus or species identification within the WMS data can be improved either by assembly of short reads or by application of longer reads (Tran and Phan 2020; Pessi et al. 2023). In some cases, workable DNA-barcode loci can be difficult to assemble from short reads due to their highly conserved sequences making the taxonomical assignment of MAGs challenging. As an example of such case, Pessi et al. (2023) reported that among 37 MAGs, obtained from 17 cyanobacterial mats from polar regions, only one included a complete sequence of 16S rRNA gene, therefore it was impossible to map most of the MAGs to a 16S rRNA sequences database. Since the step of assembly...
is not required for processing the output of long-read sequencing by ONT, it can be directly used for easy on-site taxonomical classification. The efficiency of this approach is additionally boosted by developing frameworks for rapid classification, like System for Mobile Analysis in Real-Time Environment (SMARTEn), which is implemented in Coriolis – a mobile metagenomic classification tool (Mikalsen and Zola 2023).

WMS allows investigation of the microbial species in different natural and artificially created biotopes, from natural habitats to laboratory and industrial cultures. One of the most valuable outputs of WMS of natural communities is the information about the genetic diversity of microalgae and evaluation of their physiological potential. This direction is highly contributed by large international projects aiming at collecting metagenomic samples from wide geographical area covering a lot of diverse habitats. These are represented by Tara Oceans Expedition, Microbial Atlas, etc. which have produced a large amount of data for metagenomic mining (Delmont et al. 2022). More advanced sample collection techniques, like targeting the layers of water column with the maximum chlorophyll a concentration or filtering the cells by their size, help to narrow the microbial diversity of a sample and thus improve metagenomic algorithms output (Yergeau et al. 2017; Delmont et al. 2022; Duncan et al. 2022). This enables study the genetic variability of a particular microalgal species, such as the chlorophyte Bathycoccus prasinos—a dominating member of marine eukaryotic picoplankton.

On the practical side, functional analysis of the MAGs showed amino acids content shift among polar populations of microalgae, which explains adaptation to the changes in temperatures (Duncan et al. 2022). Studying the functional landscape of eukaryotic and prokaryotic MAGs in picoplankton also allows prediction of microbiome succession, including such crucial events such as microalgal blooms (Kavagutti et al. 2023). The same approach can be used for revealing the functional potential of microalgal species discovered within metagenomes for the destruction of hazardous micropollutants by search for the relevant metabolic pathways. Examples include plastic biodegradation by adhesion on cell surface with following enzymatic hydrolysis; this process is extensively studied with the focus on the enzymes polyethylene hydrolase (PETase) and mono(2-hydroxyethyl) terephthalic acid hydrolase (MHETase) (Chia et al. 2020). Other examples include heavy metal phycoremediation by their uptake by and enzymatic reduction (e.g., by chromium reductase) in the microalgal cells (Priya et al. 2022), and xenobiotics degradation (Cheng et al. 2021; Ovis-Sánchez et al. 2023; Vasilieva et al. 2023) e.g., by nitrilase (Virgiani et al. 2019).

Though known sets of genes in metagenome can be detected by targeted PCR-analysis with degenerative primers (Gulvik et al. 2012), the results of this approach might be compromised. One of the reasons is functional redundancy—presence of alternative pathways of similar function in the community (Graham et al. 2015), another one is the functional divergence of orthologs within a species (Ma et al. 2021). Therefore, WMS remains a powerful approach for estimating the efficiency and stability of microalgal communities under particular conditions as well as for bio-monitoring of promising strains from e.g., wastewater stabilization/oxidation ponds or other polluted areas (Chia et al. 2020; Jankowski et al. 2022; Nagarajan et al. 2022). The investigation of the genetic variation landscape for microalgal and cyanobacterial species is a promising way to mine new homologs of biotechnologically valuable enzymes or alterations in biosynthetic pathways. A pangenomic analysis of databases-retrieved Nannochloropsis species genomes revealed length and sequence variations between photosystem I and II genes (psaB, J, L, and psbH, Y, N, I, T), energy conservation genes (atpH, G, E), as well as loss of the acetohydroxyacid synthase negative feedback regulation gene (ilvH) in branched chain amino acids pathway, that indicated its alternative regulation (Starkenburg et al. 2014).

Pangenomic studies have demonstrate that transposable elements are as important for the phenotype of algae as single nucleotide polymorphism (SNP), indicating the importance of sequencing method precision (Carrier et al. 2024). While deep shotgun NGS sequencing provides good nucleotide resolution, the accuracy of the assembly can be greatly enhanced by joint application of the genome-wide chromosome conformation capture (Hi-C) method with nanopore sequencing of long reads (Pan et al. 2023; Carrier et al. 2024). Being originally developed for chromatin-DNA interaction studies within a given eukaryotic species, Hi-C showed great potential in reconstruction of high-quality MAGs from microbial communities (called metaHi-C), as the capturing technique artificially gathers DNA molecules within each organism and thus improves metagenomic binning procedure (Beitel et al. 2014). The combination of short-read NGS, long-read nanopore sequencing and metaHi-C opens up the opportunity for pangenomic analysis within a certain microalga-bacterial community and the development of metapangenomic approach (Delmont and Eren 2018).

**Getting insights into interactions within microalgal-bacterial communities with HTS**

The WMS is also a powerful tool for studying the interactions within microalga-bacteria communities. On one hand, it relies on search for a specific set of genes encoding pathways for molecular signaling and/or trophic substrate exchange. Trophic relationships can be revealed starting already from elemental metabolism level, by classification.
of the relevant genes found in the MAGs as related, e.g., to phosphorus, nitrogen, or sulfur fluxes between microalgal and bacteria in a community (Saini et al. 2023). The keen attention to this kind of studies is due to importance of microalga-bacterial consortia for nutrient biosequestration from wastewater ponds, marine sediments, and biofertilizer-treated soils (Vućić and Müller 2021), where both sides can affect phosphorus accessibility for each other by enzymatic solubilization by bacteria (Dong et al. 2022) or pH modulation by the microalgae. The balance in flux between carbon, oxygen and nitrogen is crucial for the aerobic enhanced biological phosphorus removal (EBPR) process in microalgae-bacteria biofilms during wastewater treatment (Mohamed et al. 2021).

Another well-known mode of interaction between microalgae and bacteria is syntrophy, where bacterial organisms produce the vitamins biotin, cobalamin and thiamin, for which most of microalgae are auxotrophic and require them for e.g., lipid biosynthesis (Wirth et al. 2020). Comparison of metabolic potential and substrate spectrum can also reveal the spatial interaction within the cyanosphere (cyanobacterial analogue of phycosphere), where filamentous cyanobacteria (representatives of Lyngbya, Planktothrix, Pseudochroococcus and other genera) are able to build extracellular polymeric substance (EPS) of polysaccharide mucilage, which is then inhabited by heterotrophic bacteria capable of its partial degradation and utilization in catabolic reactions (Halary et al. 2022). Besides that, a more specific interaction way exists in a form of signaling molecules exchange within such consortia: phytohormones are produced by bacteria with either stimulating or suppressing mode for microalgae (for example most known L-amino oxidase manages conversion of L-tryptophan to indole-3-acetic acid) (Wang et al. 2021; Mars Brisbin et al. 2022), algicides that cause microalgal cell damage (Jia et al. 2023), and other quorum sensing with wide spectrum of impacts on photosynthetic cells (Dow 2021). One can do WMS data mining not only for the biosynthetic pathways for these operating molecules, but also for the related molecular transporters, like ABC-transporters (Krohn-Molt et al. 2017; Li et al. 2022). While solid evidence of chemical interaction between microalgae and bacteria usually requires integration with other omics (ideally proteomics and metabolomics methods), WMS provides firm background for genome-centric approach in such studies (Krohn-Molt et al. 2017).

An interesting and promising approach to investigate microalgal-bacterial interactions is one based on hologenome concept. The phycosphere can be considered as a classical holobiont—metaorganism, where certain bacteria persist and co-evolve with microalgae acting as the ecosystem engineer (edificator). That co-evolution might be revealed by comparative genomics through searching for phylosymbiotic signals (correlation in divergence) in phylogeny of both host and symbiont, codivergence of dominant microbiome groups with a host, and metabolic complementary (Cooke et al. 2019). The phycosphere is known to be highly dynamic system responding to biotic and abiotic factors and featuring the hologenome evolution mechanisms: amplification or reduction of bacterial partners, acquisition of new bacteria, and horizontal gene transfer (HGT) (Rosenberg and Zilber-Rosenberg 2018). Though HGT between eukaryotic and prokaryotic species faces many obstacles based on difference in genome structure and mechanisms, it has been shown that the gene flow from bacteria to microalgae does exist (Li et al. 2023a). It is most evident for the antibiotic resistance genes (ARG) transfer in environments with high evolutionary pressure, such as anthropogenically polluted sites, making it reasonable to propose a concept of ‘PollutantBiome’ as a special case of hologenome (Ashraf et al. 2023; Li et al. 2023a).

Investigation of the hologenome structure via comparative genomics requires low contamination values of MAGs, since presence of heterogenous reads in the final sequence leads to severe misinterpretation. Thus, long-read nanopore sequencing with the following polishing by NGS short reads is the best technique for revealing the status quo for holobiont and symbiont, as nanopore-produced long contigs reduce the probability of interspecies read contamination, while short reads increase consensus accuracy and enable analysis of SNP variants (Sauvage et al. 2019). In addition, long reads can be efficiently sorted not only by species of origin, but also by assignment to specific compartments within cells. The heteroplasmic and genetic variation of organellar genomes (nuclear, plastid, mitochondrial) of cellular endosymbionts can provide proof of gene transfer and metabolic complementarity between the microalgae holobiont and the symbionts (Sauvage et al. 2019).

Hologenome studies can be greatly enhanced by nanopore long-reads supported metaHi-C approach and opens new horizons for HGT studies, by making it possible to capture DNA–DNA interaction between host genome and mobile genetic elements (plasmids, viral loci, etc.) (Bickhart et al. 2022). The recently developed MetaCC algorithm has been shown to be a powerful tool for MAG reconstruction and plasmids search in complex microbial communities hybrid assembly of long and short reads (Du and Sun 2023). However, the holistic approach for investigation of microalga-bacteria communities currently remains underrepresented and still needs to be developed and critically reviewed.

**Advantages of long-read HTS for taxonomical profiling of microalgal-bacterial communities**

Opposite to the WMS, taxonomic profiling of microbial communities is based on amplicon sequencing of genetic barcodes, specifically determining taxonomical assignment.
of microorganisms. The variety of metabarcoding methods mainly depend on loci that are used for each particular group of organisms, with the main criteria of conservativity within the taxon and variability between taxa. Thus, the ribosomal operon is widely used for bacteria identification, since 16S and 23S rRNA genes, combined with ITS provides strain-level resolution. Recently Pushpakumara et al. (2023) have demonstrated the high potential of the 16S rRNA gene metabarcoding for analysis of microalgal-bacterial communities revealing previously unknown associations between microorganisms. The identification of eukaryotic microalgae usually requires other genetic barcodes, such as 18S rRNA gene, its ITS regions, or more specific rbcL and tufA.

Metabarcoding based on functional rbcL and tufA genes has several advantages over ribosomal loci, which are increased richness of a studied communities, and identification of haplotypes presence and microevolution via population genetic approach (Sauvage et al. 2016; Turk Dermastia et al. 2023). The second becomes available due to high resolution of identification provided by such barcodes, though it requires accurately considering possible errors and correction strategies. 16S and 23S rRNA genes are also applied for microalgae identification as plastid and mitochondrial ribosomal loci, which can be applied simultaneously to identify both components of microalgal-bacterial communities (Kezlya et al. 2023). At the same time, the presence of the plastid or mitochondrial ribosomal loci reads reduces community sample richness and affects diversity index estimation, and therefore is considered as unwelcomed contamination (Thomas et al. 2020). Both experimental techniques, such as physical removal of eukaryotic DNA (Demkina et al. 2023) and optimization of bacteria-specific primers for ribosomal operons, have been evaluated recently to obtain pure prokaryotic profiles (Thomas et al. 2020) as well as training bioinformatic classifiers on chloroplast-derived datasets, such as QIME2 naïve Bayes tool trained on PhytoREF database (Bonfantine et al. 2021).

Until recently, the DNA metabarcoding method was firmly based on short-read sequencing on the NGS platforms. Though widely spread and routine, it possesses severe drawbacks for studying the microbial communities of microalgae cultures and natural samples. The main and crucial drawback is that short read length limits taxonomical resolution. While the general rule states that ribosome small subunit rRNA gene and its ITS is required for strain identification, the NGS platforms of sequence-by-synthesis method has a limitation of maximum 300–500 b.p. (in case of pyrosequencing) and 150–300 b.p. (in case of Illumina), which allows reading of only part of the barcode. The V3-V4 regions of 16S rRNA gene is the most popular variant for microbiome profiling, though other regions, such as V2-V3 are shown to be more specific and provide higher taxa resolution (Bukin et al. 2019). Even then, the drawback lies in the interplay between resolution and richness of the community, as the increased specificity leads to the loss of particular groups of organisms. The rapid recent development of long read TGS technology enables full length barcode reading and thus removes the taxonomical resolution issues (Fig. 3) (Kerkhof et al. 2017; Portik et al. 2022). However, a one should carefully consider choice of sequencing platform for such purpose. Despite obvious advantages of long over short reads for barcode sequencing, either throughput or accuracy of sequencing itself can suffer in such race, which affects taxa identification. While PacBio can provide very accurate results at a low throughput, Oxford Nanopore products have increased throughput (especially with PromethION) but it is notorious for low accuracy of basecalling. Comparison of simulation results for different platforms showed that 50% exceed of sequencing launch capacity for Illumina over Oxford Nanopore can provide maximum accuracy of read classification and taxa identification (Pearman et al. 2020). Currently, there are many research directions of how to improve the accuracy of nanopore sequencing basecalling: by improving the technology itself through cross membrane voltage varying, by implementing other amplification strategies (such as The Rolling Circle Amplification to Concatemeric Consensus (R2C2) method), or by training basecaller models on specific datasets (Volden et al. 2018; Noakes et al. 2019; Ferguson et al. 2022). The last can be performed on species-specific datasets to improve minor taxa identification in environmental samples (Ciuffreda et al. 2021). It should be mentioned, that PacBio is considered as a useful and robust sequencing method for metabarcoding of relatively species-poor communities while targeting large regions of SSU (around 2500–3000 b.p.) of microeukaryotes (Tedersoo et al. 2018; Guedan et al. 2019).

Both experimental data and bioinformatics simulations prove that long read barcode sequences also contribute to greater richness of a studied community (Jamy et al. 2020; Lemoinne et al. 2023). Nanopore sequencing showed high potential of finding up to twice more hidden species compared to Illumina short read (Huggins et al. 2022; Lemoinne et al. 2023; Szoboszlay et al. 2023). This was shown to be especially useful for marine ecosystems, which usually possesses high richness, such as marine biofilms (Wang et al. 2022). Long read taxonomic profiling research on algal-bacterial communities of Ulva species has shown the decrease of microbiome richness but increase of relative abundance of MGPB Sulfitobacter and Roseobacter when passing from marine environmental samples to laboratory cultures (van der Loos et al. 2021). Nanopore sequencing has been demonstrated as a useful tool for investigation of the interactions within microalgal natural communities, such as harmful blooms of dinoflagellates. Sequencing of long ribosomal genes cluster cassette more than 3 kb long harboring 18S, ITS and partial 28S rRNA genes enabled identification of...
a nearly complete list of species, including the toxic microalgae *Alexandrium*, *Gonyaulax*, *Prorocentrum*, and *Linguodinium* (Hatfield et al. 2020). Studying the prokaryotic components of natural dinoflagellate communities by nanopore sequencing revealed associations between particular microalgal species and bacteria clades, such as *Alexandrium tamarense* and *Roseobacter* bacteria (Shin et al. 2018). The research authors propose that growth of *A. tamarense* can be promoted by sulfonate, which is produced by *Roseovarius* genus bacteria with Sox multienzyme complex (Shin et al. 2018).

Another issue to be kept in sight for metabarcoding is quantitative bias as a result of uneven amplification occurring for different barcode sequences (Pawluczky et al. 2015). Though targeting amplicons with conserved priming sites or application of degenerate primers slightly improves in that situation, they still cannot overcome another bias coming from various gene copy number in genomes (Krehenwinkel et al. 2017). In case of nanopore sequencing, lack of a DNA synthesis step during the sequencing step improves amplification bias for species abundance but does not remove it completely (Fig. 3) (Huggins et al. 2022). Application of optimized primers set for target barcode amplification can drastically improves PCR bias, as well as new possible selection and amplification strategies to create barcode libraries (Matsuo et al. 2021).

Fig. 3 Comparison of short read NGS and long read nanopore sequencing in application for taxonomic profiling of microbial communities. The genetic barcodes molecules from different species colored in green, orange, and blue. Nanopore sequencing technology enables reading whole unfragmented loci of genetic barcodes, also with only one PCR procedure during library preparation, thus contributing to lower amplification bias.

**Augmenting functional annotation of microalgal communities with advantageous HTS**

The biology of microalgal-bacterial consortia has a severe lack of understanding of the functional genetic landscape underlying interactions between these organisms. Even
though an emerging trend towards microalgal metagenomics enriches us with MAGs and other genomic information, we are far from its complete functional annotation and thus prediction of a role of a particular organism in a community. Classical workaround is complementing the genomic data with transcriptome—the approach successfully tested for microbial communities, including those sampled from the environment (Wang et al. 2020). This can be implemented within integrative omics pipelines and algorithms (like Galaxy) to create fully annotated metabolic networks of a particular MAG from a community (Schiml et al. 2023). Integration of metagenomics with metatranscriptomics (and full way down to other omics methods) enables investigation of complex interplay between abiotic factors (illumination, biogenic elements, etc.) and microalgal response in aquatic biomes, as well as microbial interactions within microalgal biofilms (Krohn-Molt et al. 2017; Trench-Fiol and Fink 2020). Recent advances in nanopore sequencing of both RT-PCR amplicons and direct RNA opened a way for unbiased and full-length transcripts reading for complex environmental communities, such as soil (Salzberg 2019; Poursalavati et al. 2023). Although this approach requires particular caution when handling RNA from samples of complex chemical mixtures and thus is hardly feasible in the field, it holds promise for simultaneous taxonomical identification and functional profiling of microbial communities with defined pipelines (Poursalavati et al. 2023). By accumulating a sufficient amount of accurate and complete metatranscriptomic data from known conditions the further reverse predictions of functional profile of a community can be made from similar environmental contexts and taxonomic profile only (Krinos et al. 2023).

The golden dream of microalgal communities’ researchers is an implementation of prediction algorithms based on taxonomical profile data to reveal functional potential of a community. Among the most popular are PICRUST(2) and Tax4Fun(2) whose main principle is comparison of OTU/ASV against the reference databases consisted of assembled metagenomes with functional annotation (Liu et al. 2020). Though highly reference-dependent, not taking into account the true physiological state of the cell as well as genome context, these algorithms were welcomed in studies of species interactions within a microalgal consortium including searches for potential N and/or P recovery bacteria for soil health mitigation or waste treatment (Zarezadeh et al. 2019). Besides the trophic interactions, this approach can reveal signaling cross-talk between algidical bacteria species and microphotroths (Le et al. 2022). Though not yet adjusted for these algorithms, the long read metabarcoding data produced by nanopore sequencing can dramatically improve the accuracy of such functional prediction, as species or strain-level information narrows down the functional landscape even within one taxon.

Increasing accuracy of species identification together with capability of capturing the community richness can greatly contribute development of Microbial Genome-Wide Association Studies (mGWAS) – an approach aiming for detection of genetic variants and genes responsible for specific phenotypic features (Power et al. 2017; San et al. 2020). Nanopore sequencing can provide post-GWAS fine-mapping of determined candidate loci for their further investigation and application (Magdy et al. 2020). For microalgal-bacterial consortia studies these can be genes encoding antimicrobial or algidical agents, growth-promoting factors, phytohormones, or members of biogenic element conversion cycles. At the end of the day, such “environmental GWAS” (“eGWAS”) can serve the great deal for microalga biotechnology by highlighting those genetic variants (strains) that might be useful for target process as a part of bioengineered consortia.

**Concluding remarks and outlook**

Admittedly, short-read NGS/SGS is now a mainstream platform for sequencing of genomes and transcriptomes as well as for providing support to other “omics” studies in microalgae. As such, the short-read sequencing has provided a plethora of invaluable insights into different aspects of microalgal biology, also crucial for microalgal biotechnology applications. Now we see that long-read sequencing platforms, especially nanopore-based sequencing technology, confidently enters the stage of algal research. This is especially true for metagenomic studies of microbial communities harboring microalgae as the edificator and other microorganisms contributing to the robustness, productivity, and biotechnological versatility of the whole community.

At the current level of sequencing technology evolution, both metagenomic strategies can be implemented with either short-read NGS or long-read nanopore sequencing. Still, it becomes increasingly obvious that the latter has distinct advantages that warrant its increasing application in this field (although the most fruitful approach is that employing both platforms). The number of publications dedicated to microalgal communities studied with well-established short read sequencing exponentially increased over last 15 years, as the number of the papers on microalgal genomics (Fig. 4).

The most promising directions of the metagenomic studies of microalgae include: 1) ecological monitoring of harmful microalgal blooms that cause economical and health treats to human activities; 2) mining of microalgal and/or associated bacterial strains for bioprospecting of biosynthetic pathways of valuable molecules (carotenoids, fatty acids, bioactive compounds); 3) strategies development for rational design of microalga-bacteria consortia for wastewater treatment, micropollutants biodegradation and enhanced bioproduct
production. None of these is reachable without the information about taxonomical structure and functional potential of communities, which can be easily obtained from HTS data, especially with rapid development of nanopore sequencing.

Systematic reports on nanopore-based studies of microalga metagenomes have started to emerge only recently, so one can anticipate a boom in this field in the next few years. To keep up with this trend, one should realize the tremendous potential of the long-read sequencing technologies for studies of the biology of microalgae. Therefore, it is important to highlight the benefits of the long-read sequencing for revealing taxonomic structure, genetic diversity, and functional potential of microalga-based communities for biotechnological applications. We hope that the present review makes a good step in this direction.

**Supplementary Information** The online version contains supplementary material available at [https://doi.org/10.1007/s10811-024-03267-1](https://doi.org/10.1007/s10811-024-03267-1).

**Author’s contributions** P.Z., M.Z. and A.S. wrote the main manuscript text, A.Z. and V.R. prepared the figures. All authors reviewed the manuscript.

**Funding** This work was supported by the project of the Interdisciplinary Scientific and Educational School of Moscow State University "Molecular technologies of living systems and synthetic biology" (#23-Sh04-45).

**Availability of data and materials** Not applicable.

**Declarations**

**Competing interests** A.S. is an Associate Editor for Journal of Applied Phycology and the peer-review process for this article was independently handled by another member of the journal editorial board.

**Ethical Approval** Not applicable.

**Conflict of Interest** A.S. is an Associate Editor for Journal of Applied Phycology and the peer-review process for this article was independently handled by another member of the journal editorial board.

**References**


Athanasopoulou K, Boti MA, Adamopoulos PG, Skourou PC, Scorilas A (2022) Third-generation sequencing: the spearhead towards the radical transformation of modern genomics. Life 12:30


Lane TW, Lane PD, Koh C-Y, Carney LT, Solberg OD (2013) Pond Crash forensics: microbiome analysis and field diagnostics. San-dia National Lab. (S.N.L.-CA), Livermore, CA (United States)


Lane TW, Lane PD, Koh C-Y, Carney LT, Solberg OD (2013) Pond Crash forensics: microbiome analysis and field diagnostics. San-dia National Lab. (S.N.L.-CA), Livermore, CA (United States)


Lane TW, Lane PD, Koh C-Y, Carney LT, Solberg OD (2013) Pond Crash forensics: microbiome analysis and field diagnostics. San-dia National Lab. (S.N.L.-CA), Livermore, CA (United States)


Santo EDE, Ishii M, Pinto UM, Matsudo MC, Carvalho JCMD (2022) Obtaining Bioproducts from the studies of signals and interactions between microalgae and bacteria. Microorganisms 10:2029


Szoboszlay M, Schramm L, Pinzauti D, Scerrì J, Sandionigi A, Biasso M (2023) Nanopore Isolatable over Illumina for 16S amplicon sequencing of the gut microbiota when species-level taxonomic classification, accurate estimation of richness, or focus on rare taxa is required. Microorganisms 11:


Upadhyay AK, Singh R, Singh DV, Singh L, Singh DP (2021) Microalgal consortia technology: A novel and sustainable approach of...
Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the ‘Author’s response’ area provided below

<table>
<thead>
<tr>
<th>Query</th>
<th>Details Required</th>
<th>Author’s Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQ1</td>
<td>Please check if the author group and affiliations are captured/presented correctly.</td>
<td></td>
</tr>
<tr>
<td>AQ2</td>
<td>Figure 3 contains poor quality of text inside the artwork. Please do not re-use the file that we have rejected or attempt to increase its resolution and re-save.</td>
<td></td>
</tr>
<tr>
<td>AQ3</td>
<td>Please provide complete bibliographic details of this reference Choi et al., 2021, Danish-Daniel et al., 2023, Dow, 2021, Gonzalez-Gonzalez and de-Bashan, 2023, Inskeep et al., 2013, Kang et al., 2021, Kezlya et al., 2023, Lane et al., 2013, Lemoinne et al., 2023, Ljaz and Quince, 2013, Szoboszlay et al., 2023, Trebuc et al., 2020, 2023, Trench-Fiol and Fink, 2020, Turk Dermastia et al., 2023, Vasilieva et al., 2023 and Wang et al., 2021.</td>
<td></td>
</tr>
</tbody>
</table>