Linking metagenomics to aquatic microbial ecology and biogeochemical cycles

Hans-Peter Grossart,1,2* Ramon Massana,3 Katherine D. McMahon,4 David A. Walsh5*

1Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany
2Potsdam University, Potsdam, Germany
3Institut de Ciències del Mar, CSIC, Barcelona, Catalonia, Spain
4Department of Civil and Environmental Engineering, and Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin
5Department of Biology, Concordia University, Canada

Abstract

Microbial communities are essential components of aquatic ecosystems through their contribution to food web dynamics and biogeochemical processes. Aquatic microbial diversity is immense and a general challenge is to understand how metabolism and interactions of single organisms shape microbial community dynamics and ecosystem-scale biogeochemical transformations. Metagenomic approaches have developed rapidly, and proven to be powerful in linking microbial community dynamics to biogeochemical processes. In this review, we provide an overview of metagenomic approaches, followed by a discussion on some recent insights they have provided, including those in this special issue. These include the discovery of new taxa and metabolisms in aquatic microbiomes, insights into community assembly and functional ecology as well as evolutionary processes shaping microbial genomes and microbiomes, and the influence of human activities on aquatic microbiomes. Given that metagenomics can now be considered a mature technology where data generation and descriptive analyses are relatively routine and informative, we then discuss metagenomic-enabled research avenues to further link microbial dynamics to biogeochemical processes. These include the integration of metagenomics into well-designed ecological experiments, the use of metagenomics to inform and validate metabolic and biogeochemical models, and the pressing need for ecologically relevant model organisms and simple microbial systems to better interpret the taxonomic and functional information integrated in metagenomes. These research avenues will contribute to a more mechanistic and predictive understanding of links between microbial dynamics and biogeochemical cycles. Owing to rapid climate change and human impacts on aquatic ecosystems, the urgency of such an understanding has never been greater.

Microbes are critical components of aquatic ecosystems, underpinning the provision of multiple ecosystem services (Azam and Malfatti 2007; Falkowski et al. 2008). Their communities, comprised of bacteria, archaea, microbial eukaryotes and viruses, represent nearly unfathomable levels of aquatic biodiversity (Thompson et al. 2017). A grand challenge facing the field of aquatic microbiology is our current inability to sufficiently understand how the metabolism of single organisms combined with abiotic and biotic interactions shape microbial communities and the ecosystem-scale biogeochemical transformations they mediate (Graham et al. 2016). This systems-level understanding requires an integration of research efforts in the fields of microbial ecology, evolution, and biogeochemistry. These fields are finding common ground and collaborative insights through the incorporation of genomic approaches into their scientific tool kits to better understand the underlying organismic processes and their regulation.

High throughput DNA sequencing technologies have dramatically and fundamentally advanced our ability to characterize microbial biodiversity at multiple levels of resolution, driving insight into the functional potential of ecosystems and generating new hypotheses to test. Metagenomic approaches, which we define as the production and analyses of shotgun genomic data from microbial assemblages, have been extensively adopted by aquatic scientists and are now routinely employed in studies of diverse aquatic habitats. We also acknowledge that the term metagenomics has been used to encompass single gene marker studies (Gilbert and Dupont 2010). In any case, microbiologists have coined the term “microbiome” to refer to the complete assemblage of microbes in a discrete habitat, as revealed mainly by “omics”
approaches. Thus, metagenomics has enabled us to dive deep into the water microbiome. Some 20 years after the term “metagenomics” was coined (Handelsman et al. 1998), metagenomics can now be considered a mature technology, where enormous data sets (Sunagawa et al. 2015; Almeida et al. 2019) and an extensive set of analysis tools are accessible to many researchers (Mitchell et al. 2017). Metagenomics has generated unprecedented insights, but a link to well-formulated questions and experiments is often missing. It remains challenging to integrate metagenomics into existing ecological and biogeochemical frameworks and to advance beyond descriptive science of “who’s there?” and “what are they able to do?” to a more predictive understanding based on mechanisms. The need for such understanding has never been more relevant in light of ongoing global climate change and massive human-induced alterations of aquatic ecosystems.

This special issue of Limnology & Oceanography focuses on recent insights into microbial ecology and biogeochemistry that were made possible through the application of metagenomic approaches, broadly defined. Here, we provide a brief introduction to the field and approaches used in metagenomics stating their potential, but do not focus on metagenomics methods and workflows per se, as numerous recent methodological reviews exist (Quince et al. 2017; Knight et al. 2018). We highlight some of the recent impactful metagenomics studies from marine and freshwater ecosystems, including those in this special issue. We then address some of the overarching concepts and challenges currently being discussed where metagenomics is involved. These include (1) the importance of integrating metagenomic approaches into experimental lab- and field-based investigations, particularly at the ecosystem scale; (2) the use of metagenomics data to inform and validate biogeochemical and metabolic models of aquatic microbial systems; and (3) the need for relevant model organisms and systems to allow deeper interpretation of metatranscriptomics data sets and enable more mechanistic hypothesis testing (Fig. 1). In spite of current challenges, the promise of meta-omics to transform our comprehension of aquatic microbial community structure and function is exciting and inspiring. There has never been a more thrilling time to be an aquatic microbiologist.

**A brief overview of metagenomic approaches**

The legacy of small subunit (SSU) ribosomal RNA (rRNA) based approaches

Metagenomics can trace its heritage back to 1986 when Pace and collaborators proposed that the diversity of natural microbial
communities could be investigated by directly cloning and analyzing DNA from the environment (Pace et al. 1986). Their strategy was based on sequencing the SSU rRNA gene, which at the time was the new gold standard phylogenetic marker gene used to reconcile all of cellular biodiversity into a single molecular tree of life (Woese 1987). Over the next 30 yr, this revolutionary approach revealed an immense diversity of microbial communities and showed that much of this diversity is not represented in microbial culture collections, leading to the concept of the “uncultured microbial majority” (Rappé and Giovannoni 2003).

Studies on the phylogenetic composition of microbial communities based on SSU rRNA genes or other single marker gene-based approaches are still among the most ubiquitously used high throughput DNA sequencing approaches in microbiology, and marker gene surveys continue to illuminate diversity and community structure patterns in previously underexplored aquatic environments. For example, in this issue, Cruaud et al. (2020) investigated the seasonal dynamics of river bacterial communities in relation to environmental conditions. The researchers report on the existence of a distinct ice-covered bacterial community with potentially unique metabolism compared to during the ice-free period. Further, in this issue, Huang et al. (2020) explored bacterial diversity associated with the rhizosphere of macrophytes. SSU rRNA-based investigations can provide clues about important biogeochemical roles played by aquatic microbes. Moreover, in this issue, Duret et al. (2020) investigated the eukaryotic diversity on marine particles using SSU rRNA genes and suggest an important role of heterotrophic protists in remineralization of particles at depth. Although SSU rRNA surveys continue to provide insight into microbial diversity and processes in aquatic ecosystems, with the potential to study temporal and biogeographic patterns in detail, the approach clearly misses the functional aspect of all the taxa detected. It is also limited in its ability to resolve finer-scale phylogenetic diversity that in many cases is relevant for both function and population-level evolution questions (Jaspers and Overmann 2004; Hahn and Pockl 2005; Hahn et al. 2016).

Moving beyond taxonomic composition to see what microbiomes can do

Compared to SSU rRNA-based and other single marker gene approaches, a community-wide genome-level analysis provides both taxonomic and functional inventories of microbial assemblages. Many analytical approaches for generating information on community structure and function from metagenomic data exist and their applicability depends in large part on the study goals (Quince et al. 2017). Early landmark work such as the discovery of proteorhodopsin (Beja et al. 2000) in the ocean used cloning of large genomic fragments followed by screening and comparatively laborious Sanger sequencing. Although this method continues to yield new discoveries (Weiland-Bräuer et al. 2017; Pushkarev et al. 2018), the majority of metagenomic work leverages “shotgun” sequencing, which involves the direct and presumably unbiased sequencing of small genomic DNA fragments extracted directly from microbial assemblages. The posterior analysis of metagenomes can broadly be catalogued into one of two classes: (1) assembly-independent, gene-centric approaches and (2) assembly-dependent, genome-centric approaches. Both are providing metagenomic data sets that are a valuable resource for the aquatic research community, when they are made available. Here we briefly describe the approaches used to generate these data sets and highlight their availability for those interested in gaining access. We reiterate that much more comprehensive reviews on the power and limitations of metagenomics already exist (Quince et al. 2017), including a thorough comparison between assembly independent and assembly-dependent approaches in (Dick 2018).

Assembly-independent approaches involve the generation of phylogenetic and functional profiles (using metabolic functions to describe communities) of a microbial community through direct annotation of the short reads produced through high throughput DNA sequencing technologies. Many computational tools exist to produce such profiles (Meyer et al. 2008; Huson et al. 2016), and once generated the phylogenetic and functional profiles can be statistically compared among different microbial communities (Fasching et al. 2020; Orland et al. 2020). To facilitate such comparative studies, a number of open source web applications provide automatic phylogenetic and functional analyses of metagenomes and also serve as repositories for metagenome data. Examples include MG-RAST (Meyer et al. 2008), IMG/M (Markowitz et al. 2013), and EBI MGnify (Mitchell et al. 2017).

A limitation of assembly-independent gene-centric approaches is a lack of information linking co-occurring metabolic functions to particular microbial populations or taxonomic groups. This limitation can be alleviated to some degree by genome-centric metagenomic approaches, which are rapidly becoming the standard of practice, largely due to innovations in genome assembly algorithms capable of handling the data sets emerging from current sequencing technologies. In genome-centric approaches, short sequencing reads are assembled de novo similar to whole-genome assembly of single organisms (Ayling et al. 2019). However, owing to the unique challenges associated with the high microdiversity of microbial communities and the resulting metagenome complexity, the result is potentially millions of contigs rather than complete genomes. Which contigs belong to the same genome, or even how many genomes are present, is unknown to the user. Hence the next step typically is to link contigs back to the genomes from which they are derived using contig “binning” approaches. These approaches are based on intrinsic genomic characteristics or the distribution and abundance of contigs across metagenome samples (Strous et al. 2012; Kang et al. 2015). The approach works particularly well with samples collected across space and time (Bendall et al. 2016; Coltriato et al. 2018). The output of assembly and binning workflows are metagenome-assembled genomes (MAGs). Careful curation of automatically generated MAGs is essential to avoid spurious interpretations based on cross-assembly and cross-binning of
contigs and numerous programs are available to aid the user in this process (Eren et al. 2015; Parks et al. 2015). Additional challenges associated with assembly-based approaches are difficulties in differentiating between closely related taxa, and assembly of MAGs from rare taxa. For the former challenge, emerging long-read sequencing technologies used in combination with short read technology have a high potential to improve MAG fidelity (Mantere et al. 2019).

Ultimately, and as we will highlight in the following sections, MAGs are a potent genomic resource for inferring metabolism of uncultured microbial groups, as well as for investigating aspects of community ecology, evolution, and biogeochemical processes. In addition to MAGs, advances in single-cell genomic approaches have also proven to be immensely powerful (Bowers et al. 2017; Seeleuthner et al. 2018). Single cell genomics involves the isolation of single cells by flow cytometry or other separation methods (microfluidics or single cell micromanipulator picking), followed by whole genome amplification, DNA sequencing, and assembly (Blainey 2013; Stepanauskas et al. 2017). The results are single-cell amplified genomes (SAGs). A benefit of single-cell approaches is that they lend themselves to targeted organism analysis, including rare taxa that may be missed by MAG-based analyses. Cell selection for DNA sequencing can be performed by picking single cells based on morphological or metabolic traits or by screening flow cytometry sorted cells by PCR for a particular phylogenetic or functional group of interest. Overall, MAGs and SAGs are data sets that capture the genomic content of naturally occurring populations and are proving to be an important and useful resource, particularly when used in combination to leverage each method’s strengths. MAGs seem to recover more complete genomes while SAGs provide nearly complete confidence in gene co-occurrence within an individual genome. Numerous MAG and SAG reference datasets of prokaryotes from marine (Delmont et al. 2018; Tully et al. 2018), estuarine (Alneberg et al. 2018), and freshwater systems (Linz et al. 2018) now exist and are publicly available to the research community. We note, however, the need for a centrally hosted and curated database of genomes with carefully assigned phylogenetic placement similar to the 16S rRNA based “FreshTrain” (Rohwer et al. 2018). Such a resource is currently in development for freshwater systems and is already available for specific water bodies such as the Baltic Sea (Alneberg et al. 2018).

Moving beyond functional potential to see what microbiomes are doing

Genome information for populations within aquatic communities provides extraordinarily valuable insights into the functional potential of those populations. However, this does not provide key information about which taxa and metabolisms are active in an ecosystem. An approach to move one step closer to this “holy grail” information is through community gene expression (i.e., metatranscriptomics) and protein expression (metaproteomics) analyses. Metatranscriptomic methods target the pool of expressed genes in a sample by sequencing reverse-transcribed mRNA. Transcripts can be annotated de novo without mapping to reference genomes, and this has provided our first glimpses into community-level expression (Haruta et al. 2009; Shi et al. 2011; Stewart et al. 2011; Hu et al. 2018). Metaproteomics methods involve the identification of expressed proteins by mass spectrometry analysis of peptides followed by searching peptide spectra against reference genome databases (Colatrani and Walsh 2015). Both methods are intimately linked to metagenomics and now that ample MAGs and SAGs are available, a more organism-centric approach for investigating gene and protein expression patterns is possible (Aylward et al. 2015; Tran et al. 2018). In this issue, Linz et al. (2020) nicely demonstrate that metatranscriptomics holds great promises for a more meaningful interpretation of (metagenomic content and our ability to link to biogeochemical processes at an ecosystem scale. These techniques are not without their limitations, since the results still cannot directly produce reaction rates or even truly quantitative estimates of pathway activity due to post-transcriptional regulation, other ways that organisms control fluxes through pathways, and that activities ultimately depend on the availability of substrates in the environment. Thus, targeted metabolic measures, in particular those at the single cell level (e.g., with NanoSIMS, Raman microscopy, and microautoradiography), will be powerful to use in combination.

Recent insights in aquatic microbial ecology and biogeochemistry driven by metagenomics

Metagenomics has enabled unprecedented insights into aquatic microbial diversity. Compared to now, past studies were limited in their scope due to the sequencing costs and challenges associated with data analysis, making the approach available to a restricted number of specialized laboratories. However, metagenomics can now be considered a mature and broadly accessible technology. Moving forward, a primary goal will be to leverage metagenomic data to further link biogeochemical processes with the ecological and evolutionary dynamics of microbial populations, as has been recognized in the past (Oremland et al. 2005). The quality and significance of insights obviously relies on the clarity of the question posed, the generation of testable hypotheses, appropriate experimental design and metadata collection as well as use of advanced statistical tools. In the following section, we highlight recent metagenomics-enabled studies that are moving our understanding of microbial contributions to aquatic systems forward, with focus on those studies included in this special issue.

Expanding the known phylogenetic and metabolic diversity of aquatic microbes

A classic contribution of metagenomics is the relentless discovery of unexpected phylogenetic and metabolic diversity in microbial systems. Examples include the discovery of rhodopsin-based phototrophy in the surface ocean (Beja et al.
2000), a capacity for ammonia oxidation in marine archaea (Hallam et al. 2006), and sulfur oxidation in abundant oxygen minimum zone bacteria (Walsh et al. 2009). In freshwater, a putative role of freshwater bacteria in extracellular electron transfer was based on MAG and SAG analysis (He et al. 2019). As example in this special issue, Rasigraf et al. (2020) provided a first description of microbial communities in sediments of the brackish Bothnian Sea. The researchers first employed 16S rRNA gene analysis to show strong vertical stratification of the bacterial and archaeal communities in sediments. The taxonomic profile was then used to guide a MAG-based analysis of microbial metabolism in the iron-rich sediment layer.

Recent studies continue to provide surprising, significant, and new phylogenetic and metabolic bacterial and archaeal diversity. For example, bacterial MAG reconstructions from deep aquifers and other environments has dramatically increased known Phylum level diversity by discovery of the Candidate Phylum Radiation within the bacterial domain (Brown et al. 2015; Hug et al. 2016), and a revised phylogeny of bacteria (Parks et al. 2018). Similarly, the reconstruction of archaeal MAGs from aquatic sediments lead to the discovery of the Asgard Archaea, which potentially represent the closest archaeal ancestors of eukaryotes (Zaremba-Niedzwiedzka et al. 2017; Bulzu et al. 2019). The Asgard harbor many genes previously thought to be unique to eukaryotes and their discovery has necessitated nothing less than a restructuring of the universal tree of life (Spang et al. 2015; Castelle and Banfield 2018).

In addition to revealing the expansive phylogenetic diversity of microbial communities, recent analyses of MAGs and SAGs from aquatic ecosystems continue to update our knowledge on metabolic capacities. In some cases, these studies have provided the first insights into the metabolism of uncultivated lineages, with important biogeochemical ramifications. For example, the marine SAR202 clade in the phylum Chloroflexi is an enigmatic group of bacteria described over 15 yr ago to be common throughout the world’s ocean (Morris et al. 2004), but its metabolic capacity remained unknown. Analysis of SAGs and MAGs have now implicated SAR202 in the oxidation of recalcitrant organic matter (Landry et al. 2017) and sulfur turnover (Mehrshad et al. 2018a) in the deep ocean, as well as terrestrial organic matter degradation in the Arctic Ocean (Colatramio et al. 2018). Likewise, recent studies have implicated uncultivated Chloroflexi lineages in dissolved organic nitrogen cycling, metabolism of one-carbon compounds and general carbon flow in lake hypolimnia (Denef et al. 2016; Mehrshad et al. 2018b). Such examples of descriptive discoveries that link metabolic pathways to particular phylogenetic groups generate many hypotheses, informing experiments to investigate metabolism and physiology of uncultured taxa.

Metagenomic studies have also redefined the ecological niches of specific microbial taxa and implicated them in important biogeochemical transformations. For example, marine SAR11 bacteria, which can comprise half of all bacterial cells in the oxic surface ocean (Giovannoni 2017), were recently shown to include lineages with a potential for anaerobic respiration using nitrate as electron acceptor (Tsementzi et al. 2016). In this elegant study, the nitrate reductase genes were experimentally verified to encode proteins catalyzing the first step of denitrification, using heterologous expression and activity assays. These SAR11 lineages are common in marine oxygen minimum zones, where they may play a major role in ocean nitrogen loss, and the results basically redefined the ecological niches of one of the ocean’s most abundant bacteria. Additionally, recent examples include the discovery of nitrogen fixation pathways in MAGs from surface ocean heterotrophs, including Planctomycetes (Delmont et al. 2018). In this issue, Xing et al. (2020) further implicate Planctomycetes in nitrogen cycling through the identification of nitrate reductase operons in MAGs from an alpine monomictic lake sampled during the holomictic period. With the availability of extensive MAG and SAG collections, the ecological niches of aquatic microbial assemblages will continue to be refined.

In addition to revealing metabolic features of bacteria and archaea, metagenomics and metatranscriptomics are promising tools to explore the evolution, ecology, and physiology of microbial eukaryotes through gene content (Caron et al. 2017). The activity and function of microbial eukaryotes in natural ecosystems is not based on a large flexibility of their metabolic capacities, as characteristic of prokaryotes, but on the exploration of innovations in their structural complexity (larger sizes, multicellularity, outer structures) and behaviors (motility and chemotaxis, phagocytosis, interactions; Keeling and Campo 2017). Thus, we do not expect metagenomics to reveal many fundamentally novel metabolic pathways within eukaryotes. Nevertheless, there are several eukaryotic metabolisms that can indeed be targeted through metagenomics. Examples are the search of genes involved in photosynthesis, genes for digestive enzymes that participate in phagocytosis, or genes encoding rhodopsins that can be involved in sensing the environment or act as proton pumps to produce energy or to acidify food vacuoles. In addition, fungi are being increasingly recognized as common in aquatic microbiomes and may harbor previously underexplored metabolic capabilities and genes (Grossart et al. 2019). Metagenomics can identify these genes in natural assemblages (Carradec et al. 2018), and reference genomes are then needed to assign the different versions of the same gene to a given eukaryotic species. As in prokaryotes, a large share of the microbial eukaryote diversity is uncultured, and a very successful way to recover genomic information from uncultured taxa is by single cell genomics (Yoon et al. 2011; Mangot et al. 2017). In this issue, Labarre et al. (2020) show an example of combining metatranscriptomics and single cell genomics to identify a set of genes putatively associated to phagocytosis that are actively expressed in heterotrophic uncultured microbial eukaryotes growing by bacterivory.

**Illuminating community assembly rules for aquatic microbial systems**

In addition to being critically important for ecosystem function, aquatic microbial communities are excellent model
systems for learning fundamental principles of community assembly and dynamics. In general, the field of community ecology seeks to understand the drivers and processes of community assembly that produce community structure patterns across space and time (Roughgarden 2009). Questions in this field are most often addressed from a taxonomic perspective where the distribution of species is the focus of analyses, although functional trait-based perspectives are also commonly employed (Weithoff and Beisner 2019). In microbial community ecology, the immense phylogenetic diversity of microorganisms and the relatively poor understanding of many taxon-specific traits represent considerable challenges to interpret community membership patterns and their functional implications (Krause et al. 2014; Hall et al. 2018). Functional community profiling using metagenomics is therefore useful because it can (1) simplify microbial systems to levels more amenable to statistical analysis and modeling, and (2) reveal community functional patterns across environmental gradients that may be easier to interpret from a biogeochemical perspective.

A useful conceptual framework that originated in macrobial ecology recognizes four processes in community assembly: selection, drift, dispersal, and diversification (Vellend 2010). These processes are widely recognized to be important in the assembly of microbial communities (Nemergut et al. 2013) and their strengths in shaping both taxonomic and functional composition in different microbial groups, systems, and scales can be investigated using metagenomic profiling. In this special issue, Wang et al. (2020b) compared the assembly processes shaping bacterial and archaeal communities in a subtropical river-bay system using SSU rRNA gene analyses. The authors report that deterministic (i.e., selective) processes were more pronounced than stochastic processes (i.e., drift) in shaping bacterial and archaeal communities, and that salinity and water temperature were significant drivers of community assembly. This study et al highlight that environmental drivers are important in shaping microbial community assembly (Bachmann et al. 2018).

In addition to the abiotic environment, biotic interactions are important in structuring microbial communities (Bižić-Ionescu et al. 2018). With respect to ecosystem function, the biological interactions between phytoplankton and bacteria represent one of the most important ecological relationships in aquatic environments. These interactions range from competitive to cooperative. At a basic level, the phytoplankton–bacteria relationship is based on resource provision. Bacteria can obtain a large amount of their carbon requirement directly from phytoplankton (Morán et al. 2002). On the other hand, bacteria provide limiting nutrients such as N and P, vitamins and growth factors to phytoplankton via remineralization and biosynthesis (e.g., Cole 1982). Underlying this simple scenario though are numerous known and unknown specific phytoplankton-bacteria interactions (Amin et al. 2012; Seymour et al. 2017; Zoccarato and Grossart 2019). Evidence for the selective association between bacteria and phytoplankton is the consistent detection of certain bacterial taxa in association with phytoplankton (cultures and natural blooms). This interaction through nutrient cycling explains the observed tight coupling between phytoplankton productivity and bacterial abundances at large spatial and temporal scales (Ducklow 1999). Time series studies on short temporal scales have also been important in revealing the interactions between phytoplankton and heterotrophic bacteria (Aylward et al. 2015). An example is included in this special issue by Linz et al. (2020). In this study, the researchers used metatranscriptomics to report diel changes on gene expression on three whole freshwater communities of different trophic state. They report clear diurnal transcription trends for genes related to photosynthesis, sugar transport, and carbon fixation. Photosynthesis genes are highly expressed during the day, while the expression of genes for sugar transport occurs typically a few hours later. This study demonstrates how meta-omics approaches can provide valuable mechanistic information on metabolic interactions between phototrophs and heterotrophic bacteria important to carbon cycling in aquatic ecosystems.

In recent years, there has been a growing awareness that bacteria show a wide range of beneficial interactions, even cooperative behaviors in which one population helps another at a potential cost to itself (D’Souza et al. 2018). An example is “public goods,” which are metabolites that are costly to produce yet are released into the environment and available to other populations such as iron scavenging molecules. Such interactions can involve different populations that reciprocally exchange metabolites such as sugars, amino acids, and growth factors (Garcia et al. 2018). In some cases, these synergies have been shown to emerge through direct complementation of metabolic repertoires such as amino acid auxotrophy in methanogenic consortia (Embree et al. 2015). A survey of metabolic exchange in hundreds of communities showed metabolic complementarity is common in natural communities (Zeleznikai et al. 2015). In this issue, Fernandez et al. (2020) report metagenomic evidence in support of a beneficial interaction involving metabolic exchange between nitrogen fixing bacteria and methanogenic archaea that is based on the supply of bioavailable nitrogen in exchange for carbon compounds produced by the methanogenic assemblages. Hooker et al. (2020) investigate bacterial interactions associated with Microcystis, while Wang et al. (2020a) used a combination of SSU rRNA and functional gene profiling to investigate microbial community assembly and biotic interactions that differ between macrophyte and phytoplankton dominated regimes in a shallow lake. The researchers describe a combination of complex biological interactions that differentially shape the taxonomic and functional composition of the two systems. The emergence, ecology, and evolution of microbial interactions deserve increasing attention, especially as they relate to structure–function relationships and govern biogeochemical processes.
Relationships between taxonomic and functional profiles

An additional important ecology concept that has recently been addressed using metagenomic approaches is the role of functional redundancy in shaping microbial communities. Functional redundancy refers to when the same metabolic function is present in multiple coexisting taxonomically distinct organisms (Louca et al. 2018). The case for functional redundancy is often made through comparison of taxonomic and functional profiles along spatial and temporal gradients. Such profiles are typically strongly correlated across intense transition zones, such as along redox gradients (Stewart et al. 2011) or oceanic depth profiles (DeLong 2006; Hu et al. 2018). Here, there is strong partitioning of metabolic guilds (e.g., phototrophs, obligate aerobes, denitrifiers, or methanogens) based on resource availability or thermodynamic constraints. As different taxonomic groups contribute to different metabolic guilds, structured functional profiles are expected to have underlying taxonomic profiles. This has been generally found in a recent study involving coastal microbial eukaryotes (Ramond et al. 2019). In this issue, Fasching et al. (2020) compared microbial assemblages in streams that differed in their catchments by land use type: agriculture, forested, or wetland. Consistent profile trends were observed across streams and certain functions were associated with different land use: labile DOM (agriculture), functions related to monomer uptake and carbohydrates (forest), and functions related to nitrogen metabolism and processing of aromatic carbon compounds (wetlands). These results suggest that land use influences both the taxonomic and functional composition of stream communities and hence their contributions to biogeochemical processes.

In contrast, in some systems such as the surface ocean, functional composition tends to be far more stable than taxonomic composition, suggesting functional redundancy (Louca et al. 2016b), although the degree of functional redundancy is debated (Galand et al. 2018). As an extreme example of functional redundancy, oxygenic phototrophy is common to all regions of the ocean, but the taxonomic composition of phytoplankton varies tremendously (Litchman and Klausmeier 2008). Therefore, when comparing communities, the consistency between taxonomic and functional composition depends on the relative importance of mechanisms selecting for specific functions vs. mechanisms causing substantial variation within functional groups. This idea is well illustrated by the study contributed to this special issue by Orland et al. (2020). Here the researchers experimentally investigated the influence of resource availability (terrestrial organic matter) and environmental conditions (different lakes) on the assembly of sediment microbial communities. Taxonomic dissimilarity between lakes was observed while functional profiles based on genes for terrestrial organic matter transformations were more similar, suggesting a degree of functional redundancy in the terrestrial organic matter degrading communities. Understanding the degree to which whole microbial communities, as well as specific metabolic groups, display functional redundancy is critical in assessing and predicting the biogeochemical implications of microbial dynamics in a changing environment.

In this special issue, one study demonstrates the insights that can be gleaned when analyses address factors shaping the composition within a particular functional group. Larkin et al. (2020) suggest that marine picophytoplankton (Prochlorococcus) populations (microdiverse haplotypes) are so exquisitely adapted to environmental conditions that their distributions can reveal subtle biogeochemical variability in the ocean. Overall, the distinction between the taxonomic and functional structure of the entire community, as well as the finer scale diversity within functional groups can provide valuable insights into factors shaping the assembly, composition, and ultimately function of microbial communities. This is crucial to project future changes in bacterial community composition and their related functions in a rapidly changing world.

Assessing microbial responses to climate change, land use, and pollution

Climate change is motivating much of the ecological research focused on linking microbial community dynamics to biogeochemical processes (Hutchins and Fu 2017). Although all aquatic ecosystems will be impacted to some extent, those at high latitudes such as in the Arctic are particularly vulnerable to change since warming is occurring 2–3 times more rapidly than the global average (IPCC 2007). Warming is leading to increased permafrost melt and precipitation in terrestrial environments (Vonk et al. 2012; Bintanja and Andry 2017), as well as freshening and warming of the Arctic Ocean (Carmack et al. 2016), with a myriad of potential effects on microbial ecosystem structure and biogeochemical processes (Li et al. 2009; Vincent 2010). Over the past few years, well-designed SSU rRNA and metagenomic studies are beginning to focus on understanding the taxonomic and functional composition of Arctic freshwater (Crevecoeur et al. 2015; Woodcroft et al. 2018) and marine ecosystems (Comeau et al. 2011; Yergeau et al. 2017). In some cases, these studies represent the first descriptions of important aquatic systems. For example, Colatrario et al. (2018) provided the first reference MAG data set for the Arctic Ocean, while in this issue Ruuskanen et al. (2020) generate an important MAG reference data set from sediment bacterial communities of Lake Hazen, a large high-Arctic lake. In the Lake Hazen study, they describe several genomic adaptations to life at low temperature and oligotrophic conditions. MAG data sets such as these serve as an important baseline for studying temporal changes. In a related study in this issue, Peura et al. (2020) studied the genetic potential of microbial communities to biodegrade recently mobilized permafrost carbon in 12 thermokarst ponds representing three different stages of pond succession. The authors found a clear change with pond succession in the properties of the DOM pool and in the number of carbon
degradation genes in ponds of different successional age. As change continues, metagenomic time-series data sets in combination with experimental manipulations and modeling will be essential in understanding impacts in the Arctic, as well as in other sensitive ecosystems.

In addition to global changes such as atmospheric warming, land use is an important anthropogenic impact that is having profound regional effects on coastal marine and freshwater communities and biochemistry (Camargo and Alonso 2006; Breitburg et al. 2018). As such, ecological questions on the influence of land use on aquatic microbial communities are also receiving growing attention. Several case studies incorporating metagenomics are presented in this special issue. Fasching et al. (2020) used weighted gene co-occurrence network analysis of rivers and identified specific clusters of functions related to DOM composition and land use. This study indicates distinct changes in the functional composition and loss of functional diversity of microorganisms when comparing natural to agricultural catchments. Kleinteich et al. (2020) studied the unintentional release of a large amount of ammonia nitrate into a river system, which resulted in a massive phytoplankton bloom. SSU (16S rDNA) and 18S rRNA) sequencing showed that bacterial and eukaryotic richness was reduced and community composition changed during the bloom, concluding that N pulses can have a significant effect on river communities even of naturally eutrophied water bodies. Bier et al. (2020) investigated the influence of coal mine pollutant on benthic stream communities using metagenomics, qPCR, and enzyme assays. They show that functional genes and pathways of microbial communities growing in mine effluent differ in composition, but not diversity. Thereby, the majority of functional genes and pathways that changed decreased at sites exposed to mine effluent indicating a severe effect of coal mine pollutants on microbial composition and ecosystem functions. Additional studies in this special issue related to pollution include an investigation of microbiomes associated with microbial mats in Australia (Mendes et al. 2020) as well as an investigation of marine organisms potentially involved in mercury transformation in the ocean (Bowman et al. 2020). In moving forward, a combination of environmental metagenomic surveys and experimental manipulations of pollution exposure should provide important insights into microbial responses to human influences on aquatic systems.

Evolutionary metagenomics

Metagenomics also enables investigation of evolutionary dynamics in microbial assemblages, in the context of ecological interactions. Previously, genome studies of closely related strains of bacteria and archaea have documented a high degree of genomic diversity, leading to the concept of the core and flexible genome (Cordero and Polz 2014). The core genome includes the genomic regions shared by a set of strains, while flexible genomes are comprised of regions that can vary in their distribution across strains. The ecological and functional relevance of such genomic variation has been the focus of a significant amount of recent research in aquatic microbiology. On one hand, MAGs and SAGs provide a window into the naturally occurring genome diversity of microbial populations. On the other, metagenomics provides a method to quantify the distribution and relative abundance of genomes (or parts of genomes) in the environment by fragment recruitment analyses.

Through precise tracking of populations, the ecological role of genome variation in general and the flexible genome in particular can be linked to niche partitioning of microbial genotypes. For example, Prochlorococcus, an abundant phototroph in the surface ocean has been shown to be comprised of numerous ecological populations that exhibit niche-partitioning based on light and nutrient availability (Biller et al. 2015). A study of Prochlorococcus SAGs provided evidence for hundreds of populations, differentiated by distinct genomic backbones (i.e., core genomes) linked to small sets of distinct flexible genes (Kashtan et al. 2014). Through a combination of metagenomics it was shown that environmental variation shapes the functional capacity and ecosystem role of Prochlorococcus (Kent et al. 2016). (Delmont and Eren 2018) linked metagenomes to Prochlorococcus genomes to show that closely related high-light populations exhibited variation in relative abundance linked to a few gene clusters. Similar genome ecology has been investigated for marine heterotrophs. For example, marine Alteromonas genome diversity is characterized by flexible genome regions. Interestingly, the flexible regions contain strain-specific glycosidic receptors that are exchanged by homologous recombination driven potentially by phage infection (López-Pérez and Rodriguez-Valera 2016). In freshwaters, (Bendall et al. 2016) showed selective sweeps of whole genomes and genomic regions within populations through time-series metagenomic analyses. MAGs and SAGs have also revealed unusual genome diversity in specific assemblages. For example, Ionescu et al. (2017) showed an exceptional degree of genome variation in Achromatium cells possessing multiple genome copies per cell. These studies demonstrate how systematic metagenomics analysis of microbial populations drives deeper insights into mechanisms underlying the evolutionary dynamics of microbial assemblages.

Future outlooks

Experimental manipulations to test hypotheses

Although observational studies are necessary first steps for revealing patterns in microbial biogeography and dynamics, experimental approaches are equally important as they are more amenable to hypothesis testing (Fig. 1). A large amount of metagenomic work relies on correlation-based analyses that relate one set of observations (e.g., gene profiles) to another (e.g., water chemistry), at the expense of uncovering mechanisms. On the other hand, manipulation experiments offer the potential for highly controlled mechanistic insights into relationships
between a driver and response, but perhaps at the cost of ecological relevance. Thus, the next logical step for linking microbial composition with biogeochemical processes is to design specific experiments to test hypotheses formulated based on metagenomic profiling (Krempaska et al. 2018). Recent exemplarily experimental approaches include assessment of functional redundancy along environmental gradients using mesocosms (Beier et al. 2017), microbial colonization of phytoplankton using an innovative flow-through incubation system (Bižić-lonescu et al. 2018), and nutrient limitation in microbial systems using on-ship bottle experiments (Bertrand et al. 2015). We expect that studies combining metagenomics with well-designed ecological experiments will soon provide much insight into the natural dynamics, complexity, and function of aquatic microbial communities. Moreover, such studies are becoming more feasible as sequencing costs continue to decline, allowing experimental designs to test multiple factors with suitable replication.

In this issue, a number of exciting experimental studies of biogeochemical processes that incorporate metagenomic analyses are included. In an elegant study, Bulseco et al. (2020) investigated the influence of increased nitrogen availability on salt marsh communities by coupling metagenomics with biogeochemical rate measurements. The authors show that underlying metabolic pathways can be linked to geochemical rates with important consequences for carbon and nitrogen cycling in coastal systems. Also in this special issue, Martiny et al. (2020) aimed to understand how phytoplankton adapt to differences in phosphate availability and the implications for nutrient uptake rates using a combination of transplant experiments and metagenomics to link genes with marine biogeochemistry. They show that changes in community composition and functional genes have an important effect on nutrient uptake and regulation of biogeochemistry processes. This study demonstrates how a clever combination of experiments and metagenomics provides a better understanding of the relative role of environmental conditions vs. microbial diversity in driving important ecosystem processes in pelagic systems.

**Integrating met-omics and microbial systems modeling**

Researchers working on microbiomes in many diverse ecosystems, including aquatic, often express interest in “modeling” microbial communities (Folows and Dutkiewicz 2011; Kreft et al. 2017). What do we mean when we talk about such models and how do ‘omics methods advance our ability to develop and apply them? It is important first to be clear about the purpose of microbiome modeling. Is the aim to learn more about the subcellular fluxes through metabolic pathways for individual taxa, in order to infer substrate usage preferences or reveal unaccounted for types of metabolism? Do we want to model community dynamics using ‘omics-informed traits and structured equations that capture some level of mechanism (e.g., Monod-type growth kinetics), to see how well these predictions capture our observations? For the latter, would statistical modeling with minimal explicit mechanism be sufficient? Is our goal to incorporate microbes into ecosystem-level process models to better understand carbon and nutrient cycling, or to eventually scale up to global climate models? Do we want to forecast cyanobacterial blooms? We note that many people assume that models are primarily used to forecast. However, models are in fact most useful to learn about a system of interest because they often reveal knowledge gaps when they fail to represent observed data. The primary take-away message is the need to be more explicit about our research questions before embarking on whatever kind of model (of which there are many) we choose. There is no doubt that metagenomics can drive advances in this area.

One common kind of model used to interpret ‘omics data are a conceptual metabolic reconstruction based on annotations of genes and pathways predicted to be present in a (meta)genome. These are frequently presented as a cell cartoon showing box-and-arrow pathways with a focus on central carbon and energy metabolism, nitrogen, and phosphorus transformations, and transporters used to move substrates into or out of the cell. Such diagrams are prepared manually by inspecting gene annotations, often with assistance from software that projects predicted functions onto pathway maps, such as MetaPathways (Konwar et al. 2015) and KEGG (Kanehisa et al. 2019). The metabolic reconstructions can be further investigated to identify nutrient requirements and substrates that must be acquired from the environment, using a reverse ecology technique (Hamilton et al. 2017). An often ignored, but serious issue with such modeling approaches is that compared to the millions of proteins that have been identified in genomes, only a small percentage have been functionally studied (Chang et al. 2016). Hence there is an urgent need to close the sequence-to-function gap in microbiology if we are to model metabolism based on ‘omics data (Price et al. 2018).

Studies based on SSU rRNA gene sequencing approaches often include a modeling component, often in the form of statistical approaches that correlate the relative abundance of individual taxa to each other or to abiotic environmental parameters (Fuhrman et al. 2006; Gilbert et al. 2011). Simple correlation can also be used to create models of community membership turnover (Herren and McMahon 2018). Although correlation-based models suffer from limitations (Weiss et al. 2016; Coenen and Weitz 2018; Hirano and Takemoto 2019), not least from the challenges of working with relative abundance data, they are often powerful in their simplicity and provide opportunities to generate hypotheses about taxon–taxon interactions, individual lineage lifestyles, and ecophysiology. Phytoplankton community dynamics modeling is a rich field of study, pushing the boundaries of traditional differential-equation models to include key traits (Edwards et al. 2013). However, ‘omics data is rarely, if ever, used to inform these efforts. A few researchers have tried to reproduce bacterial community dynamics using structured equations (Dam et al. 2016). The modeled population can be defined either by taxonomic affiliation or by functional guild. Although also well developed in some engineering fields (Henze et al. 2000), much potential remains for such modeling to help us learn about
aquatic microbial community assembly and dynamics. Individual-based models are also occasionally used, with potential to capture individuality (i.e., each cell potentially behaving differently; Hellweger et al. 2016). This area has only been minimally explored thus far in aquatic ecosystems.

(Meta)genome informed modeling most often seeks to recapitulate some pattern of community functional capacity to biogeochemical transformations of interest to ecosystem-scale science (e.g., respiration, denitrification, methanogenesis). Most work thus far has taken a gene-centric approach, focusing on marker genes that can serve as proxies for biochemical pathways that are expected a priori to contribute to a biogeochemical process of interest. This is particularly applicable to microbial systems characterized by strong spatial or temporal partitioning of biochemical functions such as the redox gradients associated with oxygen depletion in marine oxygen minimum zones (OMZ) or stratified lakes. As such, the first gene-centric biogeochemical model that incorporated metagenomics was developed for the Arabian Sea OMZ (Reed et al. 2014). Model simulations supported previous observations that denitrification, rather than anammox, was the dominant N-loss pathway in the Arabian Sea. In a later study, a multiomics biogeochemical model was formulated for the redoxcline of an anoxic fjord (Louca et al. 2016a). The model could reproduce the vertical profiles of biogeochemical rates of denitrification and anammox as well as their associated functional genes. Both studies also made predictions of previously unknown links between sulfur, nitrogen, and carbon cycling in OMZs. These examples show that the incorporation of meta-omics information in the formulation and validation of biogeochemical models allows one to link biogeochemical processes with microbial population dynamics.

Such integrated gene-centric modeling approaches provide insights into microbial community metabolism and may allow for prediction of nutrient and energy cycling in a rapidly changing world. However, as in all modeling approaches there are some limitations. The approach lends itself well to modeling specialized chemolitho-autotrophs, where the reactions mediated by the modeled organisms and the associate functional marker genes are known a priori. Organotrophs that reside in an environment rich in organic compounds are much more challenging to model as identifying metabolites is not straightforward, creating difficulties with regards to thermodynamic calculations and quantification of metabolites. Functional groups characterized by high functional redundancy are also much more difficult to model by such approaches. For example, photoautotrophs cannot be differentiated by their energy source alone. The ecological niches of these organisms are defined by multiple variables such as light availability, temperature preferences, and supply of inorganic nutrients. In general, these biogeochemical models do not take into consideration other ecological or physiological factors that determine the organism’s adaptive traits (e.g., cell size and structure as well as pigmentation) and consequently distribution. Ultimately, the power of these modeling approaches is limited by the biogeochemical and physiological data derived from earlier lab experiments on mainly cultured organisms and environmental data from observations. Yet, a systematic approach to studying these controlling factors in a more holistic manner are missing, but urgently needed to define ecological niches of microorganisms and project future functional responses to a changing environment.

Very few researchers have taken on the difficult task of synthesizing the understanding gained from genome-resolved approaches (MAGs and SAGs) with structured population and geochemical modeling. One well executed study tackled the distribution of key processes in the redox stratified waters of a lake (Preheim et al. 2016; Arora-Williams et al. 2018). Here, MAGs and 16S rRNA genes were recovered from a time series collected in Upper Mystic Lake, used for metabolic reconstructions with a focus on genes involved in critical pathways such as iron and sulfur cycles, and synthesized with a spatially explicit biogeochemical model. The dream of many aquatic microbial ecologists is to link this kind of organism-process coupled model into three-dimensional hydrodynamic models to most fully simulate the physical and chemical environment experienced by microbes, while allowing for feedback of microbial activities to the environment and other community members. Again, many of us can identify with this dream, but we must pause to ask: what is our question? Do we wish to know how much organism X contributes to denitrification? Do we want to know if microbes determine whether an aquatic system is a net source or sink for carbon? Do we want to learn more about our specific pet microbial group? Do we yearn to convince our nonmicrobial colleagues that microbes are important for global geochemical cycles? Perhaps all of the above.

Outside the realm of aquatic sciences, systems biologists and engineers have well developed strategies to create genome-scale metabolic models (GSMMs), usually for single pure culture organisms with some biotechnological value (Joyce and Palsson 2006). They include flux balance analysis (FBA) and metabolic flux analysis, among others. These quantitative models are powerful approaches for investigating the metabolism, physiology, and response to growth conditions for single organisms or species (Feist et al. 2009), but have only rarely been applied to multimember groups other than heavily domesticated model organisms (but see Stolyar et al. 2007; Zhuang et al. 2011). FBA are mainly used for individual (cultured) bacterial species, but recently (Garza et al. 2018) combined GSMMs of individual bacteria with relative abundance of bacteria inferred from metagenomes. This allowed them to reveal the metabolic status of a given environment. Specifically, they inferred distinct metabolomes associated with different human body environments that were consistent with their experimental data. As metagenomic analysis of whole microbial communities became feasible, the development of GSMM-like modeling approaches for quantitatively simulating microbial interactions is gaining momentum (Zuñiga et al. 2017).
Developing model organisms and systems

Ultimately our ability to interpret metagenomes and comprehend the ecological implications of microbial community composition depends on organism-centric prior knowledge of genetics, biochemistry, and ecophysiology. Hence, there is a need for future work to include model system development, with either single representative taxa or simplified enrichment cultures, to define ecophysiology and begin the more necessarily reductionist work of linking genes to functions (Fig. 1). The immensely successful combination of field ‘omics with elegant pure-culture work is best illustrated by the body of work surrounding Prochlorococcus (Bill et al. 2015). But many other abundant and cosmopolitan lineages are frustratingly resistant to isolation. This also applies to several ecologically important guilds of microbial eukaryotes, such as the colorless bacterivorous heterotrophic flagellates that are mostly composed of diverse uncultured lineages generally exhibiting a widespread marine distribution (Massana et al. 2014; Mangot et al. 2018). Several recent breakthroughs in axenic prokaryotic cultivation have occurred since the classic work targeting Pelagibacter in the SAR11 clade (Rappé et al. 2002), including isolation of the freshwater SAR11 sister clade LD12 (Candidatus Fonsibacter; Henson et al. 2018), Methylophilales members of the LD28 group (Candidatus Methylophilum; Salcher et al. 2015) and members of the recalcitrant acl lineage of actinobacteria (Kim et al. 2019). Combined cultivation and metagenomic work focused on freshwater polynucleobacter has also provided important insights into evolution and ecological principles of this ubiquitous genus (Hahn et al. 2016; Hoetzinger and Hahn 2017) Further testament to the importance of pure cultures in aquatic microbiology is the insightful research on nitrification by marine thaumarchaeota (Santoro et al. 2019). Metagenomic surveys have shown archaeal nitrification genes are widespread in the oceans (Mosier and Francis 2011). However, pure culture work has provided numerous insights into marine thaumarchaeal physiology that explain their ecological success and biogeochemical relevance, including an exceptionally high affinity for ammonia (Martens-Habbema et al. 2009), the production of the greenhouse gas nitrous oxide (Santoro et al. 2011), and other aspects of their metabolism related to their role in marine carbon and nitrogen cycling (Santoro et al. 2019). Another recent example testifying to the power of pure cultures was the discovery of dimethylsulfinio-propionate (DMSP) production in marine bacteria and the identification of genes involved in the process (Curson et al. 2017). DMSP is the most abundant organosulfur molecule on Earth and is an important nutrient and signaling molecule that was thought to be only produced by eukaryotes. The genes for DMSP production were identified through genetic analysis. The key gene (dysB) was then found to be widespread in marine metagenomes suggesting marine bacteria contribute significantly to marine DMSP production. This study exemplifies the importance of cultures to identify the genetic basis for a previously undescribed metabolism, which can then inform metagenomic studies on the distribution of a metabolic group across ecosystems.

In addition to pure isolates, the development of mixed cultures as model systems can provide insights into the structure and function of natural communities. On one hand, mixed cultures enable cultivation of ubiquitous but hard-to-cultivate microorganisms, typically because the members mutually satisfy their respective metabolic dependencies (Garcia et al. 2015; Mu et al. 2018). On the other hand, characterization of these metabolic dependencies can reveal metabolic interactions potentially occurring in the natural environment (DSouza et al. 2018). Recently, Garcia et al. (2018) cultured a mixed freshwater model community to determine auxotrophy and intrapopulation complementarity in the community’s “interactome.” Thus, by combining the relative simplicity of these model communities with bioinformatics and modeling tools, the complex nature of microbial interactions can be investigated using the relevant model communities. Another approach is to promote a given function of interest in natural assemblages through simple community manipulations. In this issue, Labarre et al. (2020) incubated a two-trophic level microbial assemblage (bacteria and their predators) in the dark to promote bacterivory and study the expressed genes related to phagocytosis in this simple set-up. Simplified, mixed cultures, thus, are ideal to explore ecology, interactions and genetic diversity of yet uncultured microorganisms. Mixed cultures in combination with a whole array of genomic tools, however, should be done in a systematic manner to better detect generalities and patterns of metabolic functions in relation to specific, simplified microbial communities, but also allow extension to more complex, natural microbial communities.

Concluding remarks

Microbial systems exhibit a vast multiscale structure, where metabolism and biological interactions of single cells drive in concert biogeochemical cycles of aquatic ecosystem with potential consequences for global processes. No doubt, the application of metagenomic approaches to investigations of aquatic microbial assemblages are allowing numerous insights into microbial community dynamics and biogeochemical processes. In moving forward, we see a number of research avenues that will contribute to our understanding of aquatic systems across multiple spatial and temporal scales. For example, ‘omics data can be linked to “big data” such as remote sensing or automatic profiler units with the outcome of relating microbial community composition and function to a whole array of environmental variables at high and broad temporal and spatial resolution (Buttigieg et al. 2018; Huot et al. 2019). Together with machine learning approaches and artificial intelligence methods, one may be able to unravel hidden patterns in the structure–function relationship of very complex microbial communities and to elucidate the related ecological consequences. Furthermore, in order to evaluate the dynamics of key environmental drivers at the relevant spatial and temporal scales, we propose to better link field studies with targeted lab and mesocosm experiments extending from
the microscale to the macro- and even global scale and from seconds to minutes, up to days, weeks, months and even years. Several other obstacles need to be addressed to further our interpretation of complex metagenomic datasets. For example, defining more environmentally relevant model organisms is needed to increase our functional understanding of aquatic microbial communities. Moreover, the functions of many proteins and even large protein families remain unknown. Although functional characterization of proteins is difficult and laborious, an increased effort for high throughput functional analysis of uncharacterized proteins is urgently required. Finally, the clever incorporation of new methods for studying cell physiology (e.g., Raman-based spectroscopy, Nano-SIMS and others) in combination with “omics” approaches will provide the required methodological basis for an improved understanding of the underlying organismic processes/mechanisms and their environmental regulation. This knowledge is urgently needed for an effective and sustainable management of aquatic ecosystems taking the omnipresent microbial legacy into account.

References

Almeida, A., A. L. Mitchell, M. Boland, S. C. Forster, G. B. Gloo, A. Tarkowska, T. D. Lawley, and R. D. Finn. 2019. A new genomic blueprint of the human gut microbiota. Nature 568: 499–504. doi:10.1038/s41586-019-0965-1

Alneberg, J., and others. 2018. BARM and BalticMicrobeDB, a reference metagenome and interface to meta-omic data for the Baltic Sea. Sci Data 5: 180146. doi:10.1038/sdata.2018.146

Amin, S. A., M. S. Parker, and E. V. Armbrust. 2012. Interactions between diatoms and bacteria. Microbiol. Mol. Biol. Rev. 76: 667–684. doi:10.1128/MMBR.00007-12

Arora-Williams, K., S. W. Olesen, B. P. Scandella, K. Delwiche, S. J. Spencer, E. M. Myers, S. Abraham, A. Sooklal, and S. P. Preheim. 2018. Dynamics of microbial populations mediating biogeochemical cycling in a freshwater lake. Microbiome 6: 165–116. doi:10.1186/s40168-018-0556-7

Ayling, M., M. D. Clark, and R. M. Leggett. 2019. New approaches for metagenome assembly with short reads. Brief. Bioinformatics 46: D726. doi:10.1093/bib/bbz2020

Aylward, F. O., J. M. Eppley, J. M. Smith, F. P. Chavez, C. A. Scholin, and E. F. DeLong. 2015. Microbial community transcriptional networks are conserved in three domains at ocean basin scales. Proc. Natl. Acad. Sci. U. S. A. 112: 5443–5448. doi:10.1073/pnas.1502883112

Azam, F., and F. Malfatti. 2007. Microbial structuring of marine ecosystems. Nat. Rev. Micro. 5: 782–791. doi:10.1038/nrmicro1747

Bachmann, J., T. Heimbach, C. Hassenrück, G. A. Kopprio, M. H. Iversen, H.-P. Grossart, and A. Gärdes. 2018. Environmental drivers of free-living vs. particle-attached bacterial community composition in the Mauritania upwelling system. Front. Microbiol. 9: 2836. doi:10.3389/fmicb.2018.02836

Beier, S., D. Shen, T. Schott, and K. Jürgens. 2017. Metatranscriptomic data reveal the effect of different community properties on multifunctional redundancy. Mol. Ecol. 26: 6813–6826. doi:10.1111/mec.14409

Beja, O., and others. 2000. Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. Science 289: 1902–1906. doi:10.1126/science.289.5486.1902

Bendall, M. L., and others. 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISMEJ. 10: 1589–1601. doi:10.1038/ismej.2015.241

Bertrand, E. M., and others. 2015. Phytoplankton-bacterial interactions mediate micronutrient colimitation at the coastal Antarctic Sea ice edge. Proc. Natl. Acad. Sci. U. S. A. 112: 9938–9943. doi:10.1073/pnas.1501615112

Bier, R., J. Wernegreen, R. Vilgalys, J. Ellis, and E. Bernhardt. 2020. Subsidized or stressed? Shifts in freshwater benthic microbial metagenomics along a gradient of alkaline coal mine drainage. Limnol. Oceanogr. 65(S1): S277–S292

Biller, S. J., P. M. Berube, D. Lindell, and S. W. Chisholm. 2015. Prochlorococcus: The structure and function of collective diversity. Nature 13: 13–27. doi:10.1038/nr9378

Bintanja, R., and O. Andry. 2017. Towards a rain-dominated Arctic. Nature Clim. Change 7: 263–267. doi:10.1038/nclimate3240

Bižić-Ionescu, M., D. Ionescu, and H.-P. Grossart. 2018. Organic particles: Heterogeneous hubs for microbial interactions in aquatic ecosystems. Front. Microbiol. 9: 2569. doi:10.3389/fmicb.2018.02569

Blainey, P. C. 2013. The future is now: Single-cell genomics of bacteria and archaea. FEMS Microbiol. Rev. 37: 407–427. doi:10.1038/ismej.2015.241

Bowers, R. M., D. F. R. Doud, and T. Woyke. 2017. Analysis of single-cell genome sequences of bacteria and archaea. Emerg. Top. Life Sci. 1: 249–255. doi:10.1042/ETLS20160028

Bowman, K., R. E. Collins, A. Agather, C. Lamborg, C. Hammerschmidt, D. Kaul, C. Dupont, G. Christensen, and D. Elias. 2020. Distribution of mercury-cycling genes in the Arctic and equatorial Pacific oceans and their relationship to mercury speciation. Limnol. Oceanogr. 65(S1): S310–S320

Breitbart, D., and others. 2018. Declining oxygen in the global ocean and coastal waters. Science 359: eaam7240. doi:10.1126/science.aam7240

Brown, C. T., and others. 2015. Unusual biology across a group comprising more than 15% of domain bacteria. Nature 523: 208–211. doi:10.1038/nature14486

Bulseco, A., J. Vineis, A. Murphy, A. Spivak, A. E. Giblin, J. Tucker, and J. Bowen. 2020. Metagenomics coupled with biogeochemical rates measurements provide evidence that nitrate addition stimulates respiration in salt marsh sediments. Limnol. Oceanogr. 65(S1): S321–S339
dynamics in microbial communities. Proc. Natl. Acad. Sci. U. S. A. 112: 15450–15455. doi:10.1073/pnas.1506034112

Eren, A. M., Ó. C. Esen, C. Quince, J. H. Vineis, H. G. Morrison, M. L. Sogin, and T. O. Delmont. 2015. Anvi'o: An advanced analysis and visualization platform for ‘omics data. PeerJ 3: e1319–e1329. doi:10.7717/peerj.1319

Falkowski, P. G., T. Fenchel, and E. F. DeLong. 2008. The microbial engines that drive Earth’s biogeochemical cycles. Science 320: 1034–1039. doi:10.1126/science.1153213

Fasching, C., and others. 2020. Linking stream microbial community function to dissolved organic matter and inorganic nutrients. Limnol. Oceanogr. 65(S1): S71–S87

Feist, A. M., M. J. Herrgard, I. Thiele, J. L. Reed, and B. Ø. Palsson. 2009. Reconstruction of biochemical networks in microorganisms. Nature 7: 129–143. doi:10.1038/nrmicro1949

Fernandez, L., S. Bertilsson, and S. Peura. 2020. Non-cyanobacterial diazotrophs dominate nitrogen-fixing communities in permafrost thaw ponds. Limnol. Oceanogr. 65 (S1): S180–S193

Folvors, M. J., and S. Dutkiewicz. 2011. Modeling diverse communities of marine microbes. Annu. Rev. Marine. Sci. 3: 427–451. doi:10.1146/annurev-marine-120709-142848

Fuhrman, J. A., I. Hewson, M. S. Schwabach, J. A. Steele, M. V. Brown, and S. Naem. 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proc. Natl. Acad. Sci. 103: 13104–13109. doi:10.1073/pnas.0602399103

Galand, P. E., O. Pereira, C. Hochart, J.-C. Auguet, and D. Debroas. 2018. A strong link between marine microbial community composition and function challenges the idea of functional redundancy. ISME J. 12: 2470–2478. doi:10.1038/s41396-018-0158-1

Garcia, S. L., M. Buck, K. D. McMahon, H.-P. Grossart, A. Eiler, and F. Warnecke. 2015. Auxotrophy and intrapopulation complementarity in the “interactome” of a cultivated freshwater model community. Mol. Ecol. 24: 4449–4459. doi:10.1111/mec.13319

Garcia, S. L., M. Buck, J. J. Hamilton, C. Wurzbacher, H.-P. Grossart, K. D. McMahon, and A. Eiler. 2018. Model communities hint at promiscuous metabolic linkages between ubiquitous free-living freshwater bacteria. mSphere 3: 257. doi:10.1128/mSphere.00202-18

Garza, D. R., M. C. van Verk, M. A. Huynen, and B. E. Dutilh. 2018. Towards predicting the environmental metabolome from metagenomics with a mechanistic model. Nat. Microbiol. 3: 456–460. doi:10.1038/s41564-018-0124-8

Gilbert, J. A., and C. L. Dupont. 2010. Microbial metagenomics: Beyond the genome. Ann. Rev. Marine Sci. 3: 347–371.

Gilbert, J. A., and others. 2011. Defining seasonal marine microbial community dynamics. ISME J. 6: 1–11. doi:10.1038/ismej.2011.107

Giovannoni, S. J. 2017. SAR11 bacteria: The most abundant plankton in the oceans. Annu. Rev. Marine. Sci. 9: 231–255. doi:10.1146/annurev-marine-010814-015934

Graham, E. B., and others. 2016. Microbes as engines of ecosystem function: When does community structure enhance predictions of ecosystem processes? Front. Microbiol. 7: 214. doi:10.3389/fmicb.2016.00214

Grossart, H.-P., S. Van den Wyngaert, M. Kagami, C. Wurzbacher, M. Cunliffe, and K. Rojas-Jimenez. 2019. Fungi in aquatic ecosystems. Nat. Microbiol. Rev. 17: 339–354. doi:10.1038/s41579-019-0175-8

Hahn, M. W., and M. Pockl. 2005. Ecosystems of planktonic actinobacteria with identical 16S rRNA genes adapted to thermal niches in temperate, subtropical, and tropical freshwater habitats. Appl. Environ. Microbiol. 71: 766–773. doi:10.1128/AEM.71.2.766-773.2005

Hahn, M. W., J. Jezberová, U. Koll, T. Sauersessig-Beck, and J. Schmidt. 2016. Complete ecological isolation and cryptic diversity in polynucleobacter bacteria not resolved by 16S rRNA gene sequences. ISME J. 10: 1642–1655. doi:10.1038/ismej.2015.237

Hall, E. K., and others. 2018. Understanding how microbiomes influence the systems they inhabit. Nat. Microbiol. 3: 977–982. doi:10.1038/s41564-018-0201-z

Hallam, S. J., T. J. Mincer, C. Schleper, C. M. Preston, K. Roberts, P. M. Richardson, and E. F. DeLong. 2006. Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. PLoS Biol. 4: e95–e17. doi:10.1371/journal.pbio.0040095

Hamilton, J. L., S. L. Garcia, B. S. Brown, B. O. Oyserman, F. Moya-Flores, S. Bertilsson, R. R. Malmstrom, K. T. Forest, and K. D. McMahon. 2017. Metabolic network analysis and metatranscriptomics reveal auxotrophies and nutrient sources of the cosmopolitan freshwater microbial lineage acl. mSystems 2: 1034. doi:10.1128/mSystems.00091-17

Handelsman, J., M. R. Rondon, S. F. Brady, J. C. Clardy, and R. M. Goodman. 1998. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. Chem. Biol. 5: R245–R249. doi:10.1016/S1074-5521(98)90108-9

Haruta, S., S. Kato, K. Yamamoto, and Y. Igarashi. 2009. Intertwined interspecies relationships: Approaches to untangle the microbial network. Environ. Microbiol. 11: 2963–2969. doi:10.1111/j.1462-2920.2009.01956.x

He, S., M. P. Lau, A. M. Linz, E. E. Roden, and K. D. McMahon. 2019. Extracellular electron transfer may be an overlooked contribution to pelagic respiration in humic-rich freshwater lakes. T.M. mSphere 4: 537. doi:10.1128/mSphere.00436-18

Hellweg, F. L., R. J. Clegg, J. R. Clark, C. M. Plugge, and J.-U. Kreft. 2016. Advancing microbial sciences by individual-based modelling, v. 14. Nature Publishing Group, p. 461–471. doi:10.1038/nrmicro.2016.62

Henson, M. W., V. C. Lanclos, B. C. Faircloth, and J. C. Thrash. 2018. Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. ISME J. 12: 1846–1860. doi:10.1038/s41396-018-0092-2
Landry, Z., B. K. Swan, G. J. Herndl, R. Stepanauskas, and S. J. Giovannoni. 2017. SAR202 genomes from the Dark Ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. MBio 8: e00413–e00417. doi:10.1128/mBio.00413-17

Larkin, A., C. Garcia, K. Ingoglia, N. Garcia, S. Baer, B. Twining, M. Lomas, and A. Martiny. 2020. Subtle biogeochemical regimes in the Indian Ocean revealed by spatial and diel frequency of Prochlorococcus haplotypes. Limnol. Oceanogr. 64.

Li, W. K. W., F. A. McLaughlin, C. Lovejoy, and E. C. Carmack. 2009. Smallest algae thrive as the Arctic Ocean freshens. Science 326: 539–539. doi:10.1126/science.1179798

Linz, A. M., S. He, S. L. R. Stevens, K. Anantharaman, R. R. Rohwer, R. R. Malmstrom, S. Bertilsson, and K. D. McMahon. 2018. Freshwater carbon and nutrient cycles revealed through reconstructed population genomes. PeerJ 6: e6075. doi:10.7717/peerj.6075

Linz, A. M., F. Aylward, S. Bertilsson, and K. McMahon. 2020. Time-series metatranscriptomes reveal conserved patterns between phototrophic and heterotrophic microbes in diverse freshwater systems. Limnol. Oceanogr. 64.

Litchman, E., and C. A. Klausmeier. 2008. Trait-based community ecology of phytoplankton. Annu. Rev. Ecol. Evol. Syst. 39: 615–639. doi:10.1146/annurev.ecolsys.39.110707.173549

López-Pérez, M., and F. Rodriguez-Valera. 2016. Pangenome evolution in the marine bacterium Alteromonas. Genome Biol. Evol. 8: 1556–1570. doi:10.1093/gbe/evw098

Louca, S., and others. 2016a. Integrating biogeochemistry with multiomic sequence information in a model oxygen minimum zone. Proc. Natl. Acad. Sci. U. S. A. 113: E5925–E5933. doi:10.1073/pnas.1602897113

Louca, S., L. W. Parfrey, and M. Doebeli. 2016b. Decoupling function and taxonomy in the global ocean microbiome. Science 353: 1272–1277. doi:10.1126/science.aaf4507

Louca, S., and others. 2018. Function and functional redundancy in microbial systems. Nat. Ecol. Evol. 2: 936–943. doi:10.1038/s41559-018-0519-1

Mangot, J.-F., and others. 2017. Accessing the genomic information of unculturable oceanic picoeukaryotes by combining multiple single cells. Sci. Rep. 7: 41498. doi:10.1038/srep41498

Mangot, J.-F., I. Forn, A. Obiol, and R. Massana. 2018. Constant abundances of ubiquitous uncultured protists in the open sea assessed by automated microscopy. Environ. Microbiol. 20: 3876–3889. doi:10.1111/1462-2920.14408

Mantere, T., S. Kersten, and A. Hoischen. 2019. Long-read sequencing emerging in medical genetics. Front. Genet. 10: 426. doi:10.3389/fgene.2019.00426

Markowitz, V. M., and others. 2013. IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic Acids Res. 42: D568–D573. doi:10.1093/nar/gkt919

Martens-Habbena, W., P. M. Berube, H. Urakawa, J. R. de la Torre, and D. A. Stahl. 2009. Ammonia oxidation kinetics determine niche separation of nitrifying archaea and bacteria. Nature 461: 976–979. doi:10.1038/nature08465

Martiny, A., L. Ustick, C. Garcia, and M. Lomas. 2020. Genomic adaptation of marine phytoplankton populations regulates phosphate uptake. Limnol. Oceanogr. 64.

Massana, R., J. del Campo, M. E. Sieracki, S. Audic, and R. Logares. 2014. Exploring the uncultured microeukaryote majority in the oceans: Reevaluation of ribogroups within stramenopiles. ISME J. 8: 854–866. doi:10.1038/ismej.2013.204

Mehrshad, M., F. Rodriguez-Valera, M. A. Amoozegar, P. López-García, and R. Ghai. 2018a. The enigmatic SAR202 cluster up close: Shedding light on a globally distributed dark ocean lineage involved in sulfur cycling. ISME J. 12: 655–668. doi:10.1038/s41396-017-0009-5

Mehrshad, M., M. M. Salcher, Y. Okazaki, S.-I. Nakano, K. Šimek, A.-S. Andrei, and R. Ghai. 2018b. Hidden in plain sight-highly abundant and diverse planktonic freshwater chloroflexi. Microbiome 6: 176–113. doi:10.1186/s40168-018-0563-8

Mendes, M. J., R. Vogwill, K. Bischoff, and D. Gleeson. 2020. Comparative metagenomics of microbial mats from hypersaline lakes at Rottnest Island (WA, Australia), advancing our understanding of the effect of mat community and functional genes on microbialite accretion. Limnol. Oceanogr. 64.

Meyer, F., and others. 2008. The metagenomics RAST server – A public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9: 386–388. doi:10.1186/1471-2105-9-386

Mitchell, A. L., and others. 2017. EBI metagenomes in 2017: Enriching the analysis of microbial communities, from sequence reads to assemblies. Nucleic Acids Res. 46: D726–D735. doi:10.1093/nar/gkx967

Morán, X. A. G., M. Estrada, J. M. Gasol, and C. Pedrós-Alió. 2002. Dissolved primary production and the strength of phytoplankton-bacterioplankton coupling in contrasting marine regions. Microb. Ecol. 44: 217–223. doi:10.1007/s00248-002-1026-z

Morris, R. M., M. S. Rappe, E. Urbach, S. A. Connolly, and S. J. Giovannoni. 2004. Prevalence of the chloroflexi-related SAR202 bacterioplankton cluster throughout the mesopelagic zone and deep ocean. Appl. Environ. Microbiol. 70: 2836–2842. doi:10.1128/AEM.70.5.2836-2842.2004

Mosier, A. C., and C. A. Francis. 2011. Determining the distribution of marine and coastal ammonia-oxidizing archaea and bacteria using a quantitative approach. Meth. Enzymol. 486: 205–221. doi:10.1016/B978-0-12-381294-0.00009-2

Mu, D.-S., Q.-Y. Liang, X.-M. Wang, D.-C. Lu, M.-J. Shi, G.-J. Chen, and Z.-J. Du. 2018. Metatranscriptomic and comparative genomic insights into resuscitation mechanisms
during enrichment culturing. Microbiome 6: 230–215. doi: 10.1186/s40168-018-0613-2

Nemerberg, D. R., and others. 2013. Patterns and processes of microbial community assembly. Microbiol. Mol. Biol. Rev. 77: 342–356. doi:10.1128/MMBR.00051-12

Oremland, R. S., D. G. Capone, J. F. Stolz, and J. Fuhrman. 2005. Whither or wither geomicrobiology in the era of “community metagenomics.” Nature Microbiol. Rev. 3: 572–578. doi:10.1038/nrmicro1182

Orland, C., k K. Yakimovich, N. Mykytczuk, N. Basiliko, and A. Tanentzap. 2020. Think global, act local: The small-scale environment mainly influences microbial community development and function in lake sediment. Limnol. Oceanogr. 64.

Pace, N. R., D. A. Stahl, D. J. Lane, and G. J. Olsen. 1986. The analysis of natural microbial populations by ribosomal RNA sequences. In K. C. Marshall [ed.], Advances in microbial ecology. Advances in microbial ecology, v. 9. Boston, MA: Springer.

Parks, D. H., M. Chuvchina, D. W. Waite, C. Rinke, A. Skarshewski, P.-A. Chaumeil, and P. Hugenholtz. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 36: 996–1004.

Parks, D. H., M. Imelfort, C. T. Skennerton, P. Hugenholtz, and G. W. Tyson. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25: 1043–1055. doi:10.1101/gr.186072.114

Peura, S., M. Wauthy, D. Simone, A. Eiler, K. Einarsdóttir, M. Rautio, and S. Bertilsson. 2020. Cycling of carbon through a microbial loop in thermokarst thaw ponds. Limnol. Oceanogr. 64.

Preheim, S. P., and others. 2016. Surveys, simulation and single-cell assays relate function and phylogeny in a lake ecosystem. Nat. Microbiol. 1: 16130. doi:10.1038/nmmicro2016.130

Price, M. N., and others. 2018. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature 557: 503–509. doi:10.1038/s41586-018-0124-0

Pushkarev, A., and others. 2018. A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. Nature 558: 595–599. doi:10.1038/s41586-018-0225-9

Quince, C., A. W. Walker, J. T. Simpson, N. J. Loman, and N. Segata. 2017. Shotgun metagenomics, from sampling to analysis. Nat. Biotechnol. 35: 833–844. doi:10.1038/nbt.3935

Ramond, P., M. Sourisseau, N. Simon, S. Romac, S. Schmitt, F. Rigaut-Jalabert, N. Henry, C. de Vargas, and R. Siano. 2019. Coupling between taxonomic and functional diversity in protistan coastal communities. Environ. Microbiol. 21: 730–749. doi:10.1111/1462-2920.14537

Rappé, M. S., and S. J. Giovannoni. 2006. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. Nature 418: 630–633. doi:10.1038/nature00917

Rasigraf, O., N. A. G. M. van Helmond, J. Frank, W. K. Lenstra, M. Egger, C. Slomp, and M. S. M. Jetten. 2020. Microbial community composition and functional potential in Bothnian Sea sediments is linked to Fe and S dynamics and the quality of organic matter. Limnol. Oceanogr. 64.

Rigaut-Jalabert, N. Henry, C. de Vargas, and R. Siano. 2019. Phytoplankton-bacteria relationships. Nat. Microbiol. 4: 17964-z.

Reed, D. C., C. K. Algar, J. A. Huber, and G. J. Dick. 2014. Gene-centric approach to integrating environmental genomics and biogeochemical models. Proc. Natl. Acad. Sci. 111: 1879–1884. doi:10.1073/pnas.1313713111

Rohwer, R. R., J. J. Hamilton, R. J. Newton, and K. D. McMahon. 2018. TaxAss: Leveraging a custom freshwater database achieves fine-scale taxonomic resolution. mSphere 3: 686. doi:10.1128/mSphere.00327-18

Roughgarden, J. 2009. Is there a general theory of community ecology? Biol Philos 24: 521–529. doi:10.1007/s10539-009-9164-z

Ruuskanen, M., G. Colby, K. St. Pierre, V. St. Louis, S. Aris-Brosou, and A. Poulan. 2020. Microbial genomes retrieved from High Arctic lake sediments encode for adaptation to cold and oligotrophic environments. Limnol. Oceanogr. 64: S233–S247. doi:10.1002/lno.11334

Salcher, M. M., S. M. Neuenschwander, T. Posch, and J. Pernthaler. 2015. The ecology of pelagic freshwater methylo trophs assessed by a high-resolution monitoring and isolation campaign. ISME J. 9: 1–12. doi:10.1038/isemj.2015.55

Santoro, A. E., C. Buchwald, M. R. McIlvin, and K. L. Casciotti. 2011. Isotopic signature of N(2)O produced by marine ammonia-oxidizing archaea. Science 333: 1282–1285. doi:10.1126/science.1208239

Santoro, A. E., R. A. Richter, and C. L. Dupont. 2019. Planktonic marine archaea. Annu. Rev. Marine. Sci. 11: 131–158. doi:10.1146/annurev-marine-121916-063141

Seelendthner, Y., S. Mondy, V. Lombard, and others. 2018. Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. Nature Communications 9: 310–10.

Seymour, J. R., S. A. Amin, J.-B. Raina, and R. Stocker. 2017. Zooming in on the phycosphere: The ecological interface for phytoplankton-bacteria relationships. Nat. Microbiol. 2: 17065. doi:10.1038/nmicrobiol.2017.65

Shi, Y., G. W. Tyson, J. M. Eppley, and E. F. DeLong. 2011. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. ISME J. 5: 999–1013. doi:10.1038/ismej.2010.189

Spang, A., and others. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature 521: 173–179. doi:10.1038/nature14447

Stepanauskas, R., and others. 2017. Improved genome recovery and integrated cell-size analyses of individual
uncultured microbial cells and viral particles. Nat. Commun. **8**: 84–10. doi:10.1038/s41467-017-00128-z

Stewart, F. J., O. Ulloa, and E. F. DeLong. 2011. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. Environ. Microbiol. **14**: 23–40. doi:10.1111/j.1462-2920.2010.02400.x

Stolyar, S., S. Van Dien, K. L. Hillesland, N. Pinel, T. J. Lie, J. A. Leigh, and D. A. Stahl. 2007. Metabolic modeling of a mutualistic microbial community. Mol. Syst. Biol. **3**: 92. doi:10.1038/msb4100131

Strous, M., B. Kraft, R. Bisdorf, and H. E. Tegetmeyer. 2012. The binning of metagenomic Contigs for microbial physiology of mixed cultures. Front. Microbio. **3**: 1–12. doi:10.3389/fmicb.2012.00410

Sunagawa, S., and others. 2015. Structure and function of the global ocean microbiome. Science **348**: 1261359–1261359. doi:10.1126/science.1261359

Thompson, L. R., and others. 2017. A communal catalogue reveals Earth’s multiscale microbial diversity. Nature **551**: 457–463. doi:10.1038/nature24621

Tran, P., A. Ramachandran, O. Khawasik, B. B. Beisner, M. Rautio, Y. Huot, and D. A. Walsh. 2018. Microbial life under ice: Metagenome diversity and in situ activity of Verrucomicrobia in seasonally ice-covered lakes. Environ. Microbiol. **20**: 2568–2584. doi:10.1111/1462-2920.14283

Tsementzi, D., and others. 2016. SAR11 bacteria linked to ocean anoxia and nitrogen loss. Nature **536**: 179–183. doi:10.1038/nature19068

Tully, B. J., E. D. Graham, and J. F. Heidelberg. 2018. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data **5**: 170203. doi:10.1038/sdata.2017.203

Vellend, M. 2010. Conceptual synthesis in community ecology. Q. Rev. Biol. **85**: 183–206. doi:10.1086/652373

Vincent, W. F. 2010. Microbial ecosystem responses to rapid climate change in the Arctic. ISME J. **4**: 1087–1090. doi:10.1038/ismej.2010.108

Vonk, J. E., and others. 2012. Activation of old carbon by erosion of coastal and subsea permafrost in Arctic Siberia. Nature **489**: 137–140. doi:10.1038/nature11392

Walsh, D. A., E. Zaikova, C. G. Howes, Y. C. Song, J. J. Wright, S. G. Tringe, P. D. Tortell, and S. J. Hallam. 2009. Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. Science **326**: 578–582. doi:10.1126/science.1175309

Wang, Y., X. Cao, J. Zeng, H. Li, D. Zhao, and Q. Wu. 2020a. Distinct shift in bacterioplankton community composition and functional gene structure between macrophyte- and phytoplankton-dominated regimes in a large shallow lake. Limnol. Oceanogr. **64**.

Wang, Y., J. Pan, J. Yang, Z. Zhou, Y. Pan, and M. Li. 2020b. Patterns and processes of free-living and particle-associated bacterioplankton and archaeaaplankton communities in a subtropical river-bay system in South China. Limnol. Oceanogr. **64**.

Weiland-Bräuer, N., D. Langfeldt, and R. A. Schmitz. 2017. Construction and screening of marine metagenomic large insert libraries. Methods Mol. Biol. **1539**: 23–42. doi:10.1007/978-1-4939-6691-2_3

Weiss, S., and others. 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. ISME J. **10**: 1669–1681. doi:10.1038/ismej.2015.235

Weithoff, G. and B. Beisner. 2019. Measures and approaches in trait-based phytoplankton community ecology – from freshwater to marine ecosystems. Frontiers in Marine Science **6**: 1–11. doi:10.3389/fmars.2019.00040

Woese, C. R. 1987. Bacterial evolution. Microbiol. Rev. **51**: 221–271.

Woodcroft, B. J., and others. 2018. Genome-centric view of carbon processing in thawing permafrost. Nature **560**: 49–54. doi:10.1038/s41586-018-0338-1

Xing, P., Y. Tao, J. Luo, L. Wang, B. Li, H. Li, and Q. Wu. 2020. Stratification of microbiomes during the holomictic period of Lake Fuxian, an alpine monomictic lake. Limnol. Oceanogr. **64**.

Yergeau, E., C. Michel, J. Tremblay, A. Niemi, T. L. King, J. Wyglinski, K. Lee, and C. W. Greer. 2017. Metagenomic survey of the taxonomic and functional microbial communities of seawater and sea ice from the Canadian Arctic. Sci. Rep. **7**: 42242. doi:10.1038/srep42242

Yoon, H. S., D. C. Price, R. Stepansauskas, V. D. Rajah, M. E. Sieracki, W. H. Wilson, E. C. Yang, S. Duffy, and D. Bhattacharya. 2011. Single-cell genomics reveals organismal interactions in uncultivated marine protists. Science **332**: 714–717. doi:10.1126/science.1203163

Zaremba-Niedzwiedzka, K., and others. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature **541**: 353–358. doi:10.1038/nature21031

Zeleznikai, A., S. Andrejev, O. Ponomarova, D. R. Mende, P. Bork, and K. R. Patil. 2015. Metabolic dependencies drive species co-occurrence in diverse microbial communities. Proc. Natl. Acad. Sci. U. S. A. **112**: 6449–6454. doi:10.1073/pnas.1421834112

Zhuang, K., M. Izallalen, P. Mouser, H. Richter, C. Risso, R. Yoon, H. S., D. C. Price, R. Stepanauskas, V. D. Rajah, M. E. Sieracki, W. H. Wilson, E. C. Yang, S. Duffy, and D. Bhattacharya. 2011. Single-cell genomics reveals organismal interactions in uncultivated marine protists. Science **332**: 714–717. doi:10.1126/science.1203163

Zoccarato, L., and H.-P. Grossart. 2019. Relationship between lifestyle and structure of bacterial communities and their functionality in aquatic systems, p. 13–52. In C. J. Hurst [ed.], The structure and function of aquatic microbial communities. Advances in environmental microbiology book series, v. 7. Cham: Springer. doi:10.1007/978-3-030-16775-2_2

Zuniga, C., L. Zaramela, and K. Zengler. 2017. Elucidation of complexity and prediction of interactions in microbial
communities. J. Microbial. Biotechnol. 10: 1500–1522. doi: 10.1111/1751-7915.12855

Acknowledgments

We would like to thank two anonymous reviewers for their helpful comments on the manuscript. DW was supported by the Canadian National Sciences and Engineering Research Council Discovery and Canada Research Chair programs. KDM acknowledges funding from the United States National Science Foundation Long-Term Ecological Research Program (NTL-LTER DEB-1440297) and the Wisconsin Alumni Research Foundation. RM was supported by the EU project SINGEK (H2020-MSCA-ITN-2015-675752) and the Spanish project ALLFLAGS (CTM2016-75083-R, MINECO). HPG was supported by the BMBF project BIBS (01LC1501G) given by the German Ministry of Education and Science (BMBF) and the DFG projects Microprime (GR 1540/23-1) and APPS (GR 1540/30-1).

Conflict of Interest

None declared.