The Journal of Eukaryotic Microbiology

ISOP Published by the International Society of Protistologists

Journal of Eukaryotic Microbiology ISSN 1066-5234

SYMPOSIUM ARTICLE

Perspectives from Ten Years of Protist Studies by High-Throughput Metabarcoding

Luciana Santoferrara[a,b], Fabien Burki[c], Sabine Filker[d], Ramiro Logares[e], Micah Dunthorn[f] & George B. McManus[b]

a Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA
b Department of Marine Sciences, University of Connecticut, Groton, CT, USA
c Department of Organismal Biology, Program in Systematic Biology, and Science for Life Laboratory, Uppsala University, Uppsala, Sweden
d Department of Molecular Ecology, University of Kaiserslautern, Kaiserslautern, Germany
e Institute of Marine Sciences (ICM), CSIC, Barcelona, Spain
f Department of Eukaryotic Microbiology, University of Duisburg-Essen, Essen, Germany

Keywords
Biodiversity; biogeography; biological interactions; ecological patterns; function; phylogenetics.

Correspondence
L. Santoferrara, University of Connecticut, 1 University Place, Stamford, CT 06901, USA
Telephone number: +1-203-251-8513;
Email: luciana.santoferrara@uconn.edu

Received: 20 December 2019; revised 29 May 2020; accepted May 29, 2020.

doi:10.1111/jeu.12813

ABSTRACT

During the last decade, high-throughput metabarcoding became routine for analyzing protistan diversity and distributions in nature. Amid a multitude of exciting findings, scientists have also identified and addressed technical and biological limitations, although problems still exist for inference of meaningful taxonomic and ecological knowledge based on short DNA sequences. Given the extensive use of this approach, it is critical to settle our understanding on its strengths and weaknesses and to synthesize up-to-date methodological and conceptual trends. This article summarizes key scientific and technical findings, and identifies current and future directions in protist research that uses metabarcoding.

High-throughput DNA sequencing of a targeted, taxonomically useful marker in environmental samples (amplicon/tag sequencing, metagenetics or, hereafter, metabarcoding; Creer et al. 2016) enables in-depth characterization of protistan communities. Metabarcoding is the community-based counterpart of DNA barcoding, which instead focuses on the identification of individual taxa (ideally at the species level) using a standardized genetic marker (Hebert et al. 2003). These complementary approaches use markers with different variability, length, and overall efficiency depending on the study aim and eukaryotic lineage, including COI (mitochondrial cytochrome oxidase *c* subunit I gene) for animals, rbcL (plastid ribulose-1,5-biphosphate carboxylase–oxygenase large subunit gene) for plants, ITS (internal transcribed spacer region of the ribosomal RNA, rRNA, operon) for fungi, and the 18S rRNA gene for protists (Pawlowski et al. 2012; Taberlet et al. 2018).

The metabarcoding approach to characterize protist communities is typically based on short hypervariable regions (< 500 base pairs) within the 18S rRNA gene, and aims at identifying and tracking taxa in the environment. Since the first publications proposing this approach for the characterization of protistan communities (Amaral-Zettler et al. 2009; Stoeck et al. 2009), many studies have used it to examine diverse environments (aquatic, terrestrial, and host-associated), locations (from shorelines to some of the most extreme sites on Earth), and scales (from local to circumglobal). Along with exciting findings included in hundreds of publications (Fig. 1), we have also learned about technical and biological limitations and have greatly improved our interpretations of the overwhelming amount of data produced. Given the wide use of metabarcoding, it is critical to settle our understanding of its strengths and limitations and to explore how to move forward methodologically and conceptually. This paper examines past

achievements, current trends, and remaining challenges in protistan metabarcoding.

WHAT HAVE WE LEARNED WITH AND ABOUT METABARCODING?

Key scientific findings

The first studies of protistan communities based on targeted (PCR-amplified) markers used Sanger-sequenced clone libraries and started to reveal organisms unseen by microscopy, either because they are too small or inconspicuous (Díez et al. 2001; Lopez-Garcia et al. 2001; Moon-van der Staay et al. 2001), or because they are too rare (Caron and Countway 2009; Doherty et al. 2007; Pedrós-Alió 2006). This trend has been magnified by high-throughput sequencing, which has made metabarcoding much more scalable and versatile (Amaral-Zettler et al. 2009; Stoeck et al. 2009). By now, many millions of protistan metabarcodes from hundreds of samples have been generated in geographically comprehensive marine surveys such as BioMarKs in European coastal waters and sediments (Massana et al. 2015) and the International Census of Marine Microbes (Amaral-Zettler et al. 2010), as well as the circumglobal expeditions of Tara Oceans (de Vargas et al. 2015) and Malaspina (Logares et al. 2020). Time series have been ongoing for more than a decade on sites of the Mediterranean coast (Giner et al. 2019). Although less common, spatially and/or temporally comprehensive surveys have also analyzed freshwater and soil environments (Debroas et al. 2017; Filker et al. 2016; Grossmann et al. 2016). Many other studies have explored protists from a wide variety of environments, including hot springs (Oliverio et al. 2018) and the mammalian gut (Parfrey et al. 2014), just to cite a few examples. In this article, we do not aim at reviewing the vast literature produced. Instead, we have identified and exemplified three groups of key findings enabled or augmented by metabarcoding.

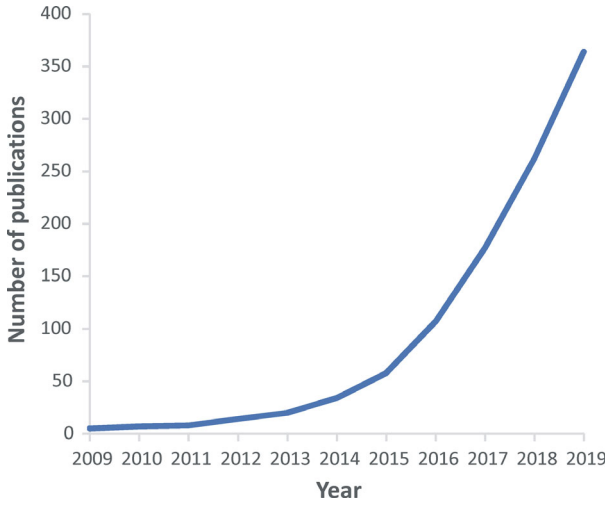

Figure 1 The use of high-throughput metabarcoding for the study of protist communities has increased exponentially in the last 10 years. Cumulative data obtained from the ISI Web of Science (accessed May 26, 2020) using the search string: ((“metabarcoding” OR “amplicon sequencing” OR “tag sequencing” OR “metagenetics” OR “eDNA” OR “environmental sequencing”) AND (“protist” OR “phytoplankton” or “microplankton” or “nanoplankton” or “picoeukaryote” OR “diatom” OR “flagellate” OR “ciliate” OR “amoeba”)).

Novel diversity and distributions

Lineages known for over two centuries turned out to be much more diverse than expected with the incorporation of metabarcoding, even in heavily studied environments such as the ocean. For example, high genetic diversity and distinct biogeographical patterns have been reported for both well-established marine groups such as dinoflagellates and diatoms (Le Bescot et al. 2015; Malviya et al. 2016) and less known clades such as MALVs and MASTs (Marine Alveolates and Marine Stramenopiles; Massana et al. 2015). Some lineages are not only unexpectedly diverse, but also unexpectedly abundant (e.g. diplonemids, syndiniales; de Vargas et al. 2015; Flegontova et al. 2016) or distributed (e.g. free-living apicomplexans; del Campo et al. 2019) in the ocean.

Dynamic communities

One common finding of metabarcoding is that communities are nearly always composed of a small number of abundant lineages accompanied by a large number of rare ones, as exemplified by rank abundance curves with very long tails (e.g. Mangot et al. 2013). This pattern is known for both macro- and microorganisms, where rare species can be either ecologically specialized or waiting for more favorable conditions to become abundant (Pedrós-Alió 2006; Preston 1948). However, the seemingly never-ending tail of rare lineages in metabarcoding studies suggests additional processes related to the large population sizes and high dispersal abilities of microbes (reviewed by Dunthorn et al. 2014a; Logares et al. 2015). Overall, metabarcoding has revealed complex combinations of ubiquitous and restricted distributions of lineages that are sometimes abundant and sometimes rare (Nolte et al. 2010; Logares et al. 2014; Ser-Giacomi et al. 2018).

Widespread interactions

In addition to supporting abiotic determinants of protistan distributions (e.g. Monier et al. 2015; Hu et al. 2016), metabarcoding data have led to new hypotheses on the crucial role of biotic interactions in shaping protistan communities. Classical knowledge on predator–prey dynamics (e.g. Montagnes et al. 2012) is now complemented with the surprising prevalence of parasitic and other symbiotic relationships across the tree of life in aquatic (e.g. Lima-Mendez et al. 2015) and terrestrial (e.g. Mahé et al. 2017) environments. Concurrent use of metabarcoding and microscope techniques such as electron microscopy and fluorescence in situ hybridization (FISH) are providing deeper knowledge on such interactions (e.g. Bird et al.

2018; Chambouvet et al. 2019; Kwong et al. 2019; Mordret et al. 2015).

Strengths and limitations of metabarcoding

After a decade of intensive use, it is easy to summarize the main advantages of metabarcoding: It allows for time- and cost-effective analysis of a large number of samples with high sensitivity and taxonomic resolution. This approach aims at providing an inventory of the lineages present in a sample and their relative abundances (or activities, if targeting RNA rather than DNA). However, like all methods, metabarcoding is prone to technical and biological limitations (reviewed by Santoferrara 2019; Taberlet et al. 2018).

Results can change markedly based on, for example, amplification primers (Hugerth et al. 2014) and regions analyzed (e.g. V4 or V9; Flegontova et al. 2016; Stoeck et al. 2010). Results can also change based on alternative bioinformatic methods that quality-filter and group sequences differently (e.g. Callahan et al. 2016; Mahé et al. 2015; Rognes et al. 2016; Schloss et al. 2009). The obtained operational taxonomic units or OTUs (also referred as to amplicon sequence variants or ASVs if obtained with methods that denoise and dereplicate sequences; e.g. Callahan et al. 2016) need careful consideration. Many of the novel or rare OTUs mentioned above could be artifacts produced during sequencing and/or sequence grouping, and thus, additional postclustering quality filters are recommended (Forster et al. 2019; Frøslev et al. 2017). Furthermore, interpretation of OTU data must consider that the relationships between species and their genomes are mostly unknown: The limits between intra- and interspecific sequence variation are usually blurry (Bachy et al. 2013; Decelle et al. 2014), and the gene copy numbers can vary by orders of magnitude even among closely related taxa (Biard et al. 2017).

Many metabarcoding issues are now well known and have been addressed, at least partially, by alternative procedures, rigorous optimization of all steps, and sequencing of replicates and controls (e.g. Decelle et al. 2014). This has resulted in an increased reliability for qualitative and some semiquantitative goals. Several studies have evaluated the metabarcoding approach by sequencing artificially assembled samples or by parallel microscope analysis, for example, based on planktonic protists (e.g. Bachy et al. 2013; Egge et al. 2013; Giner et al. 2016; Medinger et al. 2010; Santoferrara et al. 2014, 2016; Stoeck et al. 2014). These and other plankton studies have shown that sampling, laboratory and bioinformatic optimizations minimize errors such as false negatives, false positives, artifactual sequence variants and misidentifications, but relative abundances usally remain biased (Santoferrara 2019). Similarly, a higher reliability of metabarcoding for qualitative over numerical data has been shown for sediment and soil protists (Boscaro et al. 2017; Geisen et al. 2015a). Taxon disproportions can seriously distort interpretations of community structure (Medinger et al. 2010), although this is less problematic in the smallest protists (Giner et al. 2016) or for tracking the distribution of particular lineages in space and time (Pitsch et al. 2019; Santoferrara et al. 2016). Combination with other methods such as cell counting or quantitative PCR can help normalize disproportions and remains the only means to estimate absolute abundances (Canesi and Rynearson 2016; Vasselon et al. 2018; Weber and Pawlowski 2013).

Not all limitations in assessing protistan diversity and distributions can be solved with a single method or protocol. Thus, metabarcoding optimization and error management should be directed by the research question, and the selection of downstream analyses (e.g. alpha and beta diversity, and relationship with biotic and abiotic factors) should consider potentially unaddressed or remaining biases (Buttigieg and Ramette 2014; Magurran and McGill. 2011). Conclusions should also consider the nature of the parameters being measured. For example, OTUs are just a practical way to group sequences and not necessarily represent species (which are anyway difficult to define; Barraclough 2019; Boenigk et al. 2012; de Queiroz 2007). It comes as no surprise, then, that parameters such as the species richness of protists remain uncertain both locally and globally, regardless of the approach used (Caron and Hu 2019).

CURRENT AND FUTURE DIRECTIONS

There is still work to be done to make metabarcoding more accurate, reproducible, and informative. Alternative sequencing strategies, more rigorous protocols, and new bioinformatic tools continue to improve this approach. In parallel, we need to also advance conceptually and strengthen the use of metabarcoding as a complement to other techniques for inference of meaningful biological knowledge. We identify three main directions in advancing phylogenetic, ecological, and functional knowledge using metabarcoding (Fig. 2).

Improving taxonomic resolution and phylogenetic inference

The short reads typically obtained by metabarcoding contain only limited phylogenetic information, which complicates taxonomic identification (Dunthorn et al. 2014b). To handle the mass of data typically generated, it is usual to infer taxonomy based on pairwise similarity searches against a reference database (e.g. in the works by de Vargas et al. 2015; Mahé et al. 2017). While fast and relatively reliable for classification to high taxonomic ranks, similarity-based methods are heavily dependent on the taxon sampling and annotation quality of the reference databases (Berger et al. 2011). These methods also require arbitrary similarity thresholds, and because these are difficult to assess, highly divergent sequences usually remain ignored (i.e. classified as unknown, or discarded). This problem is amplified for certain environments, such as soil, because reference databases are biased toward

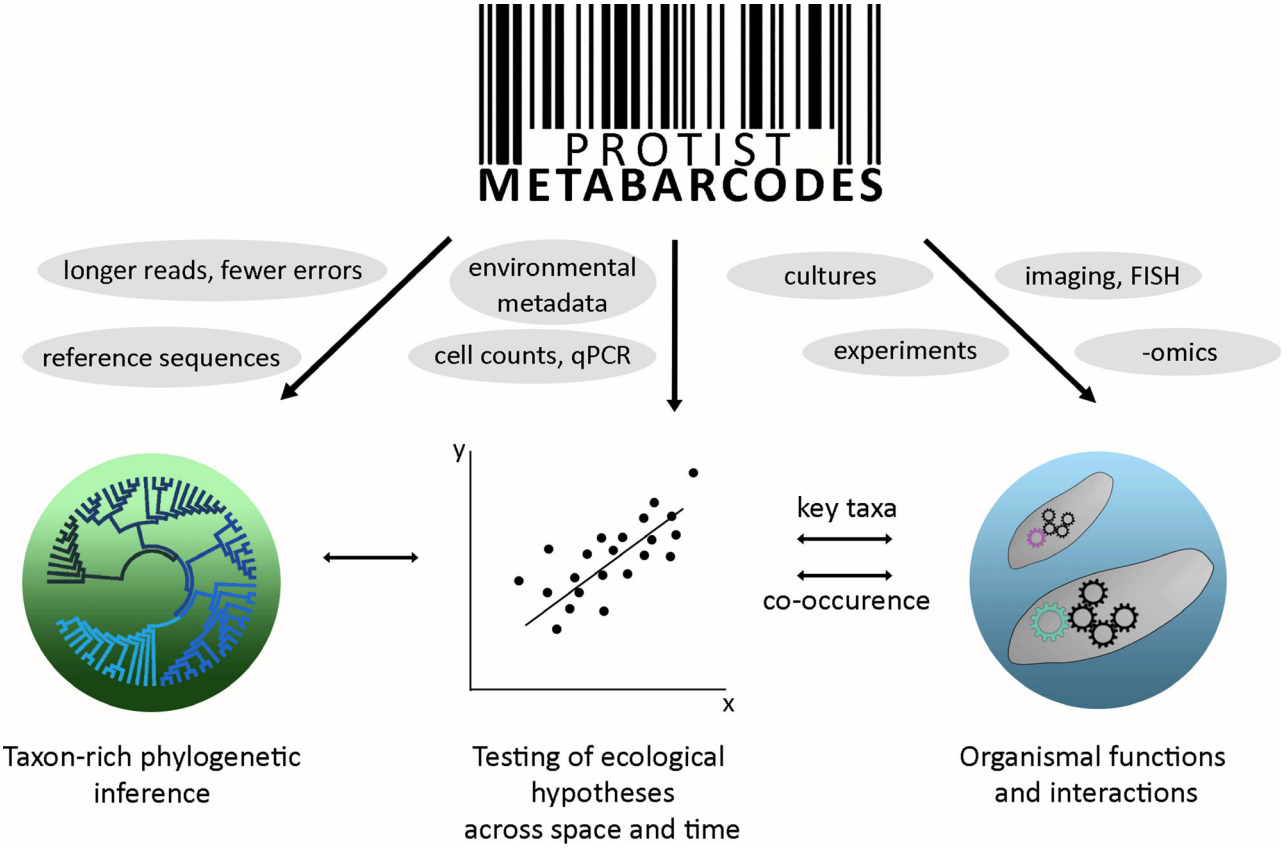

Figure 2 Methodological and conceptual advances are broadening the potential of metabarcoding data for inference of phylogenetic, ecological, and functional diversity.

aquatic organisms. As an alternative to similarity methods, phylogenetic tools (e.g. EPA and pplacer) can provide more reliable taxonomy assignments by "placing" the short reads onto a predetermined phylogeny of longer reference sequences (Barbera et al. 2019; Berger et al. 2011; Matsen et al. 2010). The major advantage of phylogenetic placement over similarity comparisons is that even sequences distantly related to references (e.g. novel groups) can be taxonomically annotated (Berger et al. 2011; Mahé et al. 2017). However, phylogenetic placement methods also present difficulties, such as the requirement of a reference phylogeny usually derived from longer but low-throughput and relatively expensive Sanger sequencing.

As a complement to traditional short-read metabarcoding, it is now possible to obtain longer sequences in a high-throughput manner from environmental samples. The Pacific Biosciences sequencer (PacBio) is particularly appealing because it allows to apply a multi-pass sequencing corrective process (circular consensus sequencing—CCS) that drastically lowers the error rates (e.g. Heeger et al. 2018). Nanopore sequencing offers other strong advantages (e.g. portability and real-time sequencing) that may make it an alternative in the near future, but due to its high error rate it is currently not best-suited to assess the diversity of complex microbial samples (Loit et al. 2019).

Using PacBio, near-full-length 18S rRNA gene sequences were obtained from environmental samples for the enigmatic protistan group of Diphyllatea (Orr et al. 2018). A few other studies mainly targeting fungal diversity were able to sequence longer fragments including also the ITS and 28S rRNA gene (Heeger et al. 2018; Tedersoo et al. 2018; Tedersoo and Anslan 2019). Most recently, Jamy et al. (2019) analyzed soil protists by producing sequences ~4,500 bp long using broad eukaryotic primers that target a region spanning from 18S to 28S rRNA genes. Taking advantage of this increased phylogenetic signal, it is possible to develop a phylogeny-aware method for taxonomic annotation. With this method, annotations are made to the appropriate taxonomic ranks corresponding to the phylogenetic position of environmental sequences in a tree that contains both reference and environmental sequences, thus allowing for a much improved taxonomically-informed evolutionary perspective on environmental DNA.

The development of long-read metabarcoding does not come without potential biases. Among the most important issues, which will all require to be properly assessed in the near future, are length differences in the target

regions that might prevent amplification of some protist groups, as well as higher risks of chimeric formation. However, depending on the study aims, some of these potential biases are outweighed by the benefits of long-read metabarcoding. One of the main advantages is that more than one molecular marker can be sequenced at a time. In the case of the rRNA operon, linked information of the 18S, ITS and 28S regions can be generated with no additional work than for "classical" metabarcodes (with the exception of optimizing the long-range PCR). Thus, widely different taxonomic resolutions can be contrasted by looking at highly variable regions (e.g. ITS) or more conserved ones (e.g. 18S rRNA gene). In addition, it is possible to populate databases for alternative markers (ITS and 28S rRNA gene) while maintaining the link to the vast body of knowledge that has accumulated for the 18S rRNA gene. Another benefit of long-amplicon sequencing, as mentioned above, is that we can produce reference phylogenies including not only reference sequences but also newly obtained environmental sequences (Jamy et al. 2019). This means that denser and more robust reference trees will become available, which, we believe, will be especially powerful for annotation of shorter Illumina sequences. Indeed, the massive throughput of Illumina means that it is possible to cover an almost unlimited range of environmental conditions and sampling strategies (e.g. different filtrations). Together with the phylogenetic signal of longer—but less abundant—PacBio or nanopore sequences, one can start sorting environmental reads at a much finer level of taxonomic resolution.

Placing metabarcodes in the context of ecological theory

Metabarcoding data have dramatically increased our estimates of microbial diversity and have provided detailed information on the composition and spatiotemporal turnover of microbial communities, especially in the ocean. Some current estimates indicate about 50,000 to 100,000 protist OTUs in the global ocean, five to ten times more than for bacteria and archaea combined (de Vargas et al. 2015; Pedrós-Alió et al. 2018). These OTUs display different distribution patterns, with diverse ocean regions typically featuring distinct microbial communities in terms of taxonomic composition and relative abundances. For example, in surface waters the distributions of prokaryotes and eukaryotes smaller than 2 μm (picoeucaryotes) seem to be determined by temperature and ocean basin, respectively (de Vargas et al. 2015; Sunagawa et al. 2015), while in the deep ocean these communities are mostly influenced by water masses (Pernice et al. 2016; Salazar et al. 2016). However, we still have a limited knowledge on the ecological mechanisms that shape microbial communities. Comprehending the underlying ecological mechanisms that determine community structures is crucial, given that these processes can lead to different ecosystem functions (Leibold et al. 2017; Mori et al. 2018; Nemergut et al. 2013). Approaches that allow not only to determine patterns, but also to use sequence data in the context of broader theory are thus crucial to answer key ecological questions.

One use of metabarcoding data to help linking distribution patterns and assembly mechanisms in microbial communities is based on null models (Stegen et al. 2013; Zhou and Ning 2017). This approach builds on community ecology theory (Vellend 2016) and quantifies the relative importance of three processes in structuring biological communities: selection (deterministic reproductive differences among individuals from different species as a response to environmental variability), dispersal (movement of species across space), and ecological drift (random changes in relative abundances derived from stochastic birth, death, immigration, and emigration); a fourth structuring process, speciation, is excluded from the estimates as it operates over evolutionary scales difficult to characterize (Stegen et al. 2013). Null models based on metabarcoding data suggest, for example, that communities of eukaryotic and prokaryotic microbes are structured by different processes in the global ocean: Picoeukaryotes are predominantly structured by dispersal limitation, while bacterial communities are shaped by the combined action of dispersal limitation, selection, and drift (Logares et al. 2020). These estimates are expected to be influenced by geographical scale (Heino et al. 2015), as exemplified by a decreased relevance of dispersal limitation in structuring both protistan and bacterial communities in surface waters of the East China Sea (Wu et al. 2018). Evolutionary processes such as local adaptation may also have an effect, as exemplified by a less important role of environmental selection than drift in shaping protistan communities (including locally adapted taxa) in lakes characterized by a strong salinity gradient (Rengefors et al. 2015).

Null models and other mathematical approaches advance our understanding of microbial communities, but should be used with caution to prevent magnifying some known issues of metabarcoding (Zhou and Ning 2017). For example, conclusions are bound to the variability of the chosen taxonomic marker. In the case of 18S rRNA gene, a small variation may reflect more than a million years of evolutionary divergence (Shapiro and Polz 2014), and thus this marker typically does not capture population level processes. Furthermore, results from this framework normally represent the action of ecological processes at the whole microbial level and not on particular taxonomic groups that can certainly be subjected to different forces. Future improvements could include algorithms to detect the action of ecological drivers at both the community and high-rank taxonomy levels, possibly using long-amplicon sequencing as outlined above. Lastly, future studies should determine whether protist communities follow additional assembly mechanisms to those known for prokaryotes, which is highly likely given fundamental differences in terms of metabolic capability, behaviors, and interactions (Massana and Logares 2013).

Linking metabarcodes with organismal functions and interactions

Knowledge on the factors and mechanisms that govern the distribution patterns of organisms is a cornerstone to better understand the processes that create and maintain biodiversity, as well as the function(ing) and dynamics of ecosystems. As discussed above, metabarcoding is an efficient tool to identify the distribution patterns of protistan taxa. These patterns can be used to develop hypotheses about underlying functions and mechanisms to be tested with additional methods.

One example where metabarcoding has served to generate hypotheses on underlying functioning refers to halotolerance in certain protists. Salinity is a major environmental determinant of community composition in bacteria (Lozupone and Knight 2007), protists (Forster et al. 2012; Logares et al. 2009), and macroorganisms (Lee and Bell 1999; Vermeij and Dudely 2000). Metabarcoding has shown specific salt transition boundaries for protists, with different physiological adaptations and osmotic capacities as the most likely explanations (Filker et al. 2017). To explore these possibilities, ciliate cultures, ion imaging, and proton nuclear magnetic resonance spectroscopy were used to investigate two main haloadaptive strategies in ciliates, that is, intracellular accumulation of inorganic ions and accumulation/synthesis of compatible solutes as osmoprotectants (Weinisch et al. 2018a,b). In agreement with the metabarcode-based transition boundaries, experimentation indicated that ciliates exhibit different salt tolerances and that all the tested species use compatible solutes to combat osmotic stress. Compatible solute concentrations also showed a linear increase with increasing external salinity, while the proportions of compatible solutes within each ciliate pointed at slight differences in haloadaptive strategies by regulatory actions (Weinisch et al. 2018a,b). These and other studies that identify novel osmoprotectant chemicals (e.g. Harding et al. 2016) serve as a basis for genome, transcriptome, and proteome analyses to elucidate underlying functional pathways.

In addition to environmental factors, complex networks of trophic interactions (e.g. predator–prey interactions, competition, and symbiosis; Bjorbækmo et al. 2019) impact the dynamics of protistan communities. Metabarcoding data (abundance- or incidence-based) in combination with correlation methods can serve as phylogenetic proxies to infer these types of interactions. Association network inferences can identify sets of microorganisms that show significant co-presences or co-absences across samples, and can also incorporate environmental trait data for prediction of species–environment relationships (Faust and Raes 2012). More complex ecological interactions, in which one species is affected (in dependence or influenced) by multiple other species, can be assessed by regression- or rule-based networks (Faust and Raes 2012). Microbial network inferences have manifold strengths, such as the integration of metabarcodes, genes, pathways, and other data types, as well as the identification of community properties encoded in the network structure (e.g. identification of keystone species). However, to be able to fully and correctly exploit these networks, one has to consider the several associated statistical pitfalls, which can heavily affect the results (Freilich et al. 2018; Röttjers and Faust 2018). In any case, information inferred from metabarcoding data can serve as starting point for detailed investigations of microbial interactions, for example, in co-culture experiments (Harcombe 2010) or using stable isotope probing, FISH, and other imaging techniques (Fig. 3; Filker et al. 2014; Kwong et al. 2019; Orsi et al. 2012, 2018).

METABARCODING OR META'OMICS?

The phase of exploration and method validation that metabarcoding underwent during the past decade is now accelerating for protistan metagenomics and metatranscriptomics (targeting full genomes or transcribed genes in an environmental sample, respectively). The critical question arises: Why would we focus on a single marker, if we can now target hundreds of genes or transcripts? The answer depends on the study aim, target lineage, and budget.

Metagenomics and metatranscriptomics are common for studying the taxonomic and functional diversity of bacteria and archaea (e.g. Salazar et al. 2019; Yooseph et al. 2010). For protists, applications have included reconstruction of full genomes or targeted protein-coding genes (Cuvelier et al. 2010; Karin et al. 2019; West et al. 2018) and estimation of gene expression in natural communities (e.g. Alexander et al. 2015; Carradec et al. 2018; Geisen

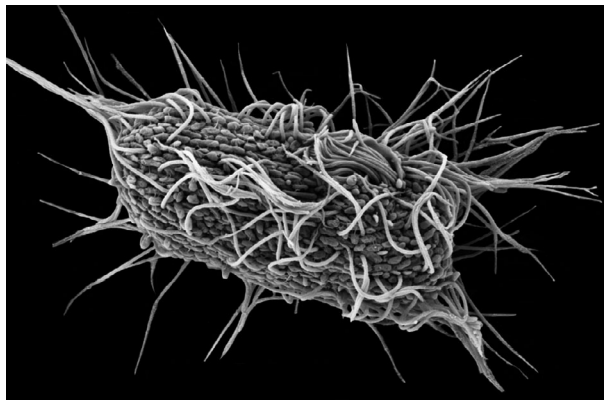

Figure 3 An example of biotic interaction that can be tracked across hypersaline environments using metabarcoding. Complementary isolation, culturing, and experimentation enabled characterization of the halophile ciliate *Platynematum salinarum* and its archaeal ectosymbiont, Candidatus *Haloectosymbiotes riaformosensis* (image obtained with scanning electron microscopy). Preliminary experiments suggest an obligate symbiosis between the two species, in which the ciliate might serve as means of transport or nutrient supplier for the ectosymbiont, in exchange for specific compatible solutes that help the ciliate to counteract high external osmotic pressure (Filker et al. 2014).

et al. 2015b). As much as these approaches are appealing, however, broad use in protistology is currently limited by a new set of challenges. Protistan genomes and transcriptomes are much more complex than prokaryotic ones. Except for protists with highly reduced genomes (e.g. certain parasites), eukaryotic genomes are usually much bigger and much richer in noncoding regions as compared to prokaryotes (Lynch and Conery 2003). While noncoding regions may be useful for certain aims (e.g. population genetics), a huge sequencing depth is needed to capture functional genes. Metatranscriptomics solves this issue by focusing on expressed functional genes, but the transient nature of messenger RNA molecules makes difficult their recovery in certain settings (Edgcomb et al. 2014). Furthermore, both metagenomics and metatranscriptomics are currently limited by the scarcity of reference protistan genomes and transcriptomes (Keeling and del Campo 2017; Sibbald and Archibald 2017), despite significant efforts such as the Marine Microbial Eukaryote Transcriptome Sequencing Project (Caron et al. 2017; Keeling et al. 2014).

Metabarcoding and meta'omics approaches address different questions, and should be seen as complementary rather than mutually exclusive. Metabarcoding remains advantageous for certain aims, for example, to document and monitor biodiversity (Taberlet et al. 2018). No other method currently allows for the analysis of large numbers of samples with the same sensitivity, taxonomic resolution, and costs. However, as much as metabarcoding has changed our view of taxonomic diversity, there is no doubt that -omics approaches (genomics, transcriptomics, proteomics, and metabolomics) used at different levels (from single cells to full communities) will change our perspectives on the functional diversity of protists.

CONCLUSION

Metabarcoding has greatly changed our view of protistan taxonomic diversity during the last decade. Still, there are remaining challenges that we should address as a scientific community. For example, we are producing large amounts of data that are hardly comparable among studies and challenge current meta-analysis efforts (e.g. EukBank, a key piece of UniEuk; Berney et al. 2017). Although aim-oriented optimization is needed to ensure accuracy within a given study, a certain degree of standardization for sampling, laboratory, and bioinformatic procedures should be achieved, for example, among studies with shared goals (e.g. biomonitoring), sample types (e.g. soil or water), or lineages. Broader, sustained sharing of sequencing data (pre- and postprocessing) and protocols would facilitate comparisons, increase reproducibility, and foster discovery. We also need continued efforts to maintain comprehensive and publicly available reference databases of taxonomically identified sequences, as these are the necessary backbone to link metabarcodes with current taxonomic frameworks (del Campo et al. 2018; Guillou et al. 2013).

As much as we saw advances in taxonomic diversity, we now need to move forward into better-resolved phylogenetic and functional diversity. Longer sequence reads are already generating better-resolved taxonomic assignments and phylogenetic inferences. Large datasets offer opportunities to explore diversity in ecological or evolutionary contexts, based on classic theory and using sound study design and statistics for hypothesis-driven science.

ACKNOWLEDGMENTS

This article resulted from the symposium "Ten Years of Metabarcoding: What Have we Learned and How Do we Move Forward?" held during the VIII European Congress of Protistology—ISOP Joint Meeting, Rome, Italy, in July 2019. We thank the meeting organizers for logistic support, and the International Society of Protistologists and the Gordon and Betty Moore Foundation for funding this symposium. Additional support for this work includes a grant by the US National Science Foundation (OCE1924527) to GM and LS, and a fellowship from the Science for Life Laboratory (SciLifeLab) to FB.

LITERATURE CITED

Alexander, H., Jenkins, B. D., Rynearson, T. A. & Dyhrman, S. 2015. Metatranscriptome analyses indicate resource partitioning between diatoms in the field. *Proc. Natl Acad. Sci. USA*, 112: E2182–E2190.

Amaral-Zettler, L., Artigas, L. F., Baross, J., Bharathi, L., Boetius, A., Chandramohan, D., Herndl, G., Kogure, K., Neal, P., Pedros-Alio, C., Ramette, A., Schouten, S., Stal, L., Thessen, A., de, Leeuw, J. & Sogin, M. 2010. A global census of marine microbes. In: McIntyre (ed.), Life in the World's Oceans: Diversity, Distribution and Abundance. Blackwell Publishing Ltd, Oxford, pp. 223–245.

Amaral-Zettler, L., McCliment, E., Ducklow, H. & Huse, S. 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE*, 4:e6372.

Bachy, C., Dolan, J., Lopez-Garcia, P., Deschamps, P. & Moreira, D. 2013. Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. *ISME J.*, 7:244–255.

Barbera, P., Kozlov, A., Czech, L., Morel, B., Darriba, D., Flouri, T. & Stamatakis, A. 2019. EPA-ng: massively parallel evolutionary placement of genetic sequences. *System. Biol.*:syy054. https://doi.org/10.1093/sysbio/sy

Barraclough, T. G. 2019: The Evolutionary Biology of Species. Oxford University Press, Oxford.

Berger, S. A., Krompass, D. & Stamatakis, A. 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *System. Biol.*, 60:291–302.

Berney, C., Ciuprina, A., Bender, S., Brodie, J., Edgcomb, V., Kim, E., Rajan, J., Parfrey, L. W., Adl, S., Audic, S., Bass, D., Caron, D. A., Cochrane, G., Czech, L., Dunthorn, M., Geisen, S., Glöckner, F. O., Mahé, F., Quast, C., Kaye, J. Z., Simpson, A. G. B., Stamatakis, A., del Campo, J., Yilmaz, P. & de Vargas, C. 2017. UniEuk: time to speak a common language in protistology!. *J. Eukaryot. Microbiol.*, 64:407–411.

Biard, T., Bigeard, E., Audic, S., Poulain, J., Gutierrez-Rodriguez, A., Pesant, S., Stemmann, L. & Not, F. 2017. Biogeography and diversity of Collodaria (Radiolaria) in the global ocean. *ISME J.*, 11:1331–1344.

Bird, C., Darling, K., Russell, A., Fehrenbacher, J., Davis, C., Free, A. & Ngwenya, B. 2018. 16S rRNA gene metabarcoding and TEM reveals different ecological strategies within the genus *Neogloboquadrina* (planktonic foraminifer). *PLoS ONE*, 13: e0191653.

Bjorbækmo, M. F. M., Evenstad, A., Røsæg, L. L., Krabberød, A. K. & Logares, R. 2019. The planktonic protist interactome: where do we stand after a century of research? *ISME J.*. https://doi.org/10.1038/s41396-019

Boenigk, J., Ereshefsky, M., Hoef-Emden, K., Mallet, J. & Bass, D. 2012. Concepts in protistology: species definitions and boundaries. *Eur. J. Protistol.*, 48:96–102.

Boscaro, V., Rossi, A., Vannini, C., Verni, F., Fokin, S. I. & Petroni, G. 2017. Strengths and biases of high-throughput sequencing data in the characterization of freshwater ciliate microbiomes. *Microb. Ecol.*, 73:865–875.

Buttigieg, P. & Ramette, A. 2014. A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. *FEMS Microbiol. Ecol.*, 90:543–550.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A. & Holmes, S. P. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Meth.*, 13:581.

del Campo, J., Heger, T. J., Rodríguez-Martínez, R., Worden, A. Z., Richards, T. A., Massana, R. & Keeling, P. J. 2019. Assessing the diversity and distribution of apicomplexans in host and free-living environments using high-throughput amplicon data and a phylogenetically informed reference framework. *Front. Microbiol.*. https://doi.org/10.3389/fmicb

del Campo, J., Kolisko, M., Boscaro, V., Santoferrara, L. F., Nenarokov, S., Massana, R., Guillou, L., Simpson, A., et al. 2018. EukRef: phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. *PLoS Biol.*, 16:e2005849.

Canesi, K. L. & Rynearson, T. A. 2016. Temporal variation of *Skeletonema* community composition from a long-term time series in Narragansett Bay identified using high-throughput DNA sequencing. *Mar. Ecol. Prog. Ser.*, 556:1–16.

Caron, D. A., Alexander, H., Allen, A. E., Archibald, J. M., Armbrust, E. V., Bachy, C., Bell, C. J., Bharti, A., Dyhrman, S. T., Guida, S. M., Heidelberg, K. B., Kaye, J. Z., Metzner, J., Smith, S. R. & Worden, A. Z. 2017. Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat. Rev. Microbiol.*, 15:6–20.

Caron, D. A. & Countway, P. D. 2009. Hypotheses on the role of the protistan rare biosphere in a changing world. *Aquat. Microb. Ecol.*, 57:227–238.

Caron, D. & Hu, S. 2019. Are we overestimating protistan diversity in nature? *Trends Microbiol.*, 27:197–205.

Carradec, Q., Pelletier, E., da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., Lima-Mendez, G., Rocha, F., et al. 2018. A global ocean atlas of eukaryotic genes. *Nat. Commun.*, 9:373.

Chambouvet, A., Monier, A., Maguire, F., Itoïz, S., del Campo, J., Elies, P., Edvardsen, B., Eikreim, W. & Richards, T. A. 2019. Intracellular infection of diverse diatoms by an evolutionary distinct relative of the fungi. *Curr. Biol.*, 29:4093–4101.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W., Potter, C. & Bik, H. 2016. The ecologist's field guide to sequence based identification of biodiversity. *Meth. Ecol. Evol.*, 7:1008–1018.

Cuvelier, M. L., Allen, A. E., Monier, A., McCrow, J. P., Messié, M., Tringe, S. G., Woyke, T., Welsh, R. M., et al. 2010. Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc. Natl Acad. Sci.*, 107:14679–14684.

Debroas, D., Domaizon, I., Humbert, J.-F., Jardillier, L., Lepère, C., Oudart, A. & Taïb, N. 2017. Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available metabarcoding data. *FEMS Microbiol. Ecol.*, 93:fix023–fix023.

Decelle, J., Romac, S., Sasaki, E., Not, F. & Mahé, F. 2014. Intracellular diversity of the v4 and v9 regions of the 18S rRNA in marine protists (radiolarians) assessed by high-throughput sequencing. *PLoS ONE*, 9:e104297.

Díez, B., Pedrós-Alió, C. & Massana, R. 2001. Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl. Environ. Microbiol.*, 67:2932.

Doherty, M., Costas, B., McManus, G. & Katz, L. 2007. Culture-independent assessment of planktonic ciliates in coastal northwest Atlantic waters. *Aquat. Microb. Ecol.*, 48:141–154.

Dunthorn, M., Otto, J., Berger, S. A., Stamatakis, A., Mahé, F., Romac, S., de Vargas, C., Audic, S., Consortium, B., Stock, A., Kauff, F. & Stoeck, T. 2014b. Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. *Mol. Biol. Evol.*, 31:993–1009.

Dunthorn, M., Stoeck, T., Clamp, J., Warren, A. & Mahé, F. 2014a. Ciliates and the rare biosphere: a review. *J. Eukaryot. Microbiol.*, 61:404–409.

Edgcomb, V. P., Taylor, C., Pachiadaki, M., Engstrom, I. & Yakimov, M. 2014. Comparison of Niskin vs. in situ approaches for analysis of gene expression in deep Mediterranean Sea water samples. *Deep Sea Res. II*, 129:213–222.

Egge, E., Bittner, L., Andersen, T., Audic, S., de Vargas, C. & Edvardsen, B. 2013. 454 pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. *PLoS ONE*, 8: e74371.

Faust, K. & Raes, J. 2012. Microbial interactions: from networks to models. *Nat. Rev.*, 10:538–550.

Filker, S., Forster, D., Weinisch, L., Mora-Ruiz, M., González, B., Farías, M., Rosselló-Móra, R. & Stoeck, T. 2017. Transition boundaries for protistan species turnover in hypersaline waters of different biogeographic regions. *Environ. Microbiol.*, 19:3186–3200.

Filker, S., Kaiser, M., Rossello-Mora, R., Dunthorn, M., Lax, G. & Stoeck, T. 2014. "Candidatus *Haloectosymbiotes riaformosensis*" (Halobacteriaceae), an archaeal ectosymbiont of the hypersaline ciliate *Platynematum salinarum*. *Syst. Appl. Microbiol.*. https://doi.org/10.1016/j.syapm.2014.01

Filker, S., Sommaruga, R., Vila, I. & Stoeck, T. 2016. Protistan plankton communities of high-mountain lakes from three continents exhibit strong biogeographic patterns. *Mol. Ecol.*. https://doi.org/10.1111/mec.13633

Flegontova, O., Flegontov, P., Malviya, S., Audic, S., Wincker, P., de Vargas, C., Bowler, C., Lukeš, J. & Horák, A. 2016. Extreme diversity of diplonemid eukaryotes in the ocean. *Curr. Biol.*, 26:3060–3065.

Forster, D., Behnke, A. & Stoeck, T. 2012. Meta-analyses of environmental sequence data identify anoxia and salinity as parameters shaping ciliate communities. *Syst. Biodivers.*, 10:277–288.

Forster, D., Lentendu, G., Filker, S., Dubois, E., Wilding, T. A. & Stoeck, T. 2019. Improving eDNA-based protist diversity assessments using networks of amplicon sequence variants. *Environ. Microbiol.*, 21:4109–4124.

Freilich, M. A., Wieters, E., Broitman, B. R., Marquet, P. A. & Navarrete, S. A. 2018. Species co-occurrence networks: can they reveal trophic and non-trophic interactions in ecological communities?. *Ecology*, 99:690–699.

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C. & Hansen, A. J. 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nat. Commun.*, 8:1188.

Geisen, S., Laros, I., Vizcaíno, A., Bonkowski, M. & de Groot, G. A. 2015a. Not all are free-living: high-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. *Mol. Ecol.*, 24:4556–4569.

Geisen, S., Tveit, A. T., Clark, I. M., Richter, A., Svenning, M. M., Bonkowski, M. & Urich, T. 2015b. Metatranscriptomic census of active protists in soils. *ISME J.*, 9:2178–2190.

Giner, C. R., Balagué, V., Krabberød, A. K., Ferrera, I., Reñé, A., Garcés, E., Gasol, J. M., Logares, R. & Massana, R. 2019. Quantifying long-term recurrence in planktonic microbial eukaryotes. *Mol. Ecol.*, 28:923–935.

Giner, C., Forn, I., Romac, S., Logares, R., de Vargas, C. & Massana, R. 2016. Environmental sequencing provides reasonable estimates of the relative abundance of specific picoeukaryotes. *Appl. Environ. Microbiol.*, 82:4757–4766.

Grossmann, L., Jensen, M., Heider, D., Jost, S., Glücksman, E., Hartikainen, H., Mahamdallie, S. S., Gardner, M., Bass, D. & Boenigk, J. 2016. Protistan community analysis: key findings of a large-scale molecular sampling. *ISME J.*, 10:2269–2279.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., et al. 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.*, 41:D597–D604.

Harcombe, W. 2010. Novel cooperation experimentally evolved between species. *Evolution*, 64:2166–2172.

Harding, T., Brown, M. W., Simpson, A. G. & Roger, A. J. 2016. Osmoadaptative strategy and its molecular signature in obligately halophilic heterotrophic protists. *Genome Biol. Evol.*, 8:2241–2258.

Hebert, P. D. N., Cywinska, A., Ball, S. & deWaard, J. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B Biol.*, 270:313–321.

Heeger, F., Bourne, E. C., Baschien, C., Yurkov, A., Bunk, B., Spröer, C., Overmann, J., Mazzoni, C. J. & Monaghan, M. T. 2018. Long-read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments. *Mol. Ecol. Resour.*, 18:1500–1514.

Heino, J., Melo, A. S., Siqueira, T., Soininen, J., Valanko, S. & Bini, L. M. 2015. Metacommunity organization, spatial extent and dispersal in aquatic systems: patterns, processes and prospects. *Freshwa. Biol.*, 60:845–869.

Hu, S. K., Campbell, V., Connell, P., Gellene, A. G., Liu, Z., Terrado, R. & Caron, D. A. 2016. Protistan diversity and activity inferred from RNA and DNA at a coastal ocean site in the eastern North Pacific. *FEMS Microbiol. Ecol.*, 92:fiw050.

Hugerth, L. W., Muller, E., Hu, Y., Lebrun, L., Roume, H., et al. 2014. Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. *PLoS ONE*, 9: e95567. https://doi.org/10.1371/journal.pone.0095567

Jamy, M., Foster, R., Barbera, P., Czech, L., Kozlov, A., Stamatakis, A., Bending, G., Hilton, S., Bass, D. & Burki, F. 2019. Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. *Mol. Ecol. Resour.*.

Karin, E. L., Mirdita, M. & Söding, J. 2019. MetaEuk—sensitive, high-throughput gene discovery and annotation for large-scale eukaryotic metagenomics. *bioRxiv*:851964. https://doi.org/10.1101/851964

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., Armbrust, E. V., Archibald, J. M., Bharti, A. K., Bell, C. J., Beszteri, B., Bidle, K. D., Cameron, C. T., Campbell, L., Caron, D. A., Cattolico, R. A., Collier, J. L., Coyne, K., Davy, S. K., Deschamps, P., Dyhrman, S. T., Edvardsen, B., Gates, R. D., Gobler, C. J., Greenwood, S. J., Guida, S. M., Jacobi, J. L., Jakobsen, K. S., James, E. R., Jenkins, B., John, U., Johnson, M. D., Juhl, A. R., Kamp, A., Katz, L. A., Kiene, R., Kudryavtsev, A., Leander, B. S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A. M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M. A., Murray, S., Nadathur, G., Nagai, S., Ngam, P. B., Palenik, B., Pawlowsky, J., Petroni, G., Piganeau, G., Posewitz, M. C., Rengefors, K., Romano, G., Rumpho, M. E, Rynearson, T., Schilling, K. B., Schroeder, D. C., Simpson, A. G. B., Slamovits, C. H., Smith, D. R., Smith, G. J., Smith, S. R., Sosik, H. M., Stief, P., Theriot, E., Twary, S. N., Umale, P. E., Vaulot, D., Wawrik, B., Wheeler, G. L., Wilson, W. H., Xu, Y., Zingone, A. & Worden, A. Z. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.*, 12:e1001889.

Keeling, P. J. & del Campo, J. 2017. Marine protists are not just big bacteria. *Curr. Biol.*, 27:R541–R549.

Kwong, W. K., del Campo, J., Mathur, V., Vermeij, M. & Keeling, P. J. 2019. A widespread coralinfecting apicomplexan with chlorophyll biosynthesis genes. *Nature*, 568:103–107.

Le Bescot, N., Mahé, F., Audic, S., Dimier, C., Garet, M.-J., Poulain, J., Wincker, P., de Vargas, C. & Siano, R. 2015. Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding. *Environ. Microbiol.*, 18: 609–626.

Lee, C. E. & Bell, M. A. 1999. Causes and consequences of recent freshwater invasions by saltwater animals. *Trends Ecol. Evol.*, 14:282–288.

Leibold, M. A., Chase, J. & Ernest, S. K. 2017. Community assembly and the functioning of ecosystems: how metacommunity processes alter ecosystems attributes. *Ecology*, 98: 909–919.

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J. C., et al. 2015. Determinants of community structure in the global plankton interactome. *Science*, 348:1262073.

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., Claverie, J., Decelle, J., Dolan, J., Dunthorn, M., et al. 2014. Patterns of rare and abundant marine microbial eukaryotes. *Curr. Biol.*, 24:813–821.

Logares, R., Bråte, J., Bertilsson, S., Clasen, J., Shalchian-Tabrizi, K. & Rengefors, K. 2009. Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol.*, 17:414–422.

Logares, R., Deutschmann, I. M., Junger, P. C., Giner, C. R., Krabberød, A. K., Schmidt, T. S. B., Rubinat-Ripoll, L., Mestre, M., Salazar, G., Ruiz-González, C., Sebastián, M., de Vargas, C., Acinas, S. G., Duarte, C. M., Gasol, J. M. & Massana, R. 2020. Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome*. https://doi.org/10.1186/s40168-020-00827-8

Logares, R., Mangot, J.-F. & Massana, R. 2015. Rarity in aquatic microbes: placing protists on the map. *Res. Microbiol.*, 166:831–841.

Loit, K., Adamson, K., Bahram, M., Puusepp, R., Anslan, S., Kiiker, R., Drenkhan, R. & Tedersoo, L. 2019. Relative performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) Third-Generation sequencing instruments in identification of agricultural and forest fungal pathogens. *Appl. Environ. Microb.*, 85:e01368-19.

Lopez-Garcia, P., Rodriguez-Valera, F., Pedros-Alio, C. & Moreira, D. 2001. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature*, 409:603–607.

Lozupone, C. & Knight, R. 2007. Global patterns in bacterial diversity. *Proc. Natl Acad. Sci. USA*, 104:11436–11440.

Lynch, M. & Conery, J. 2003. The origins of genome complexity. *Science*, 302:1401–1404.

Magurran, A. E. & McGill., B. 2011: Biological Diversity. Oxford University Press, Oxford.

Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., Singer, D., Mayor, J., Bunge, J., Sernaker, S., Siemenmeyer, T., et al. 2017. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat. Ecol. Evol.*, 1:0091.

Mahé, F., Rognes, T., Quince, C., de Vargas, C. & Dunthorn, M. 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ*, 3:e1420.

Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., de Vargas, C., Bittner, L., Zingone, A. & Bowler, C. 2016. Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl Acad. Sci. USA*, 113:E1516–E1525.

Mangot, J.-. F., Domaizon, I., Taib, N., Marouni, N., Duffaud, E., Bronner, G. & Debroas, D. 2013. Short-term dynamics of small eukaryotes. *Environ. Microbiol.*, 15:1745–1758.

Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., Chambouvet, A., Christen, R., Claverie, J.-M., Decelle, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Forn, I., Forster, D., Guillou, L., Jaillon, O., Kooistra, W. H. C. F., Logares, R., Mahé, F., Not, F., Ogata, H., Pawlowski, J., Pernice, M. C., Probert, I., Romac, S., Richards, T., Santini, S., Shalchian-Tabrizi, K., Siano, R., Simon, N., Stoeck, T., Vaulot, D., Zingone, A. & de Vargas, C. 2015. Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environ. Microbiol.*, 17:4035–4049.

Massana, R. & Logares, R. 2013. Eukaryotic versus prokaryotic marine picoplankton ecology. *Environ. Microbiol.*, 15:1254–1261.

Matsen, F. A., Kodner, R. B. & Armbrust, V. E. 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinf.*, 11:538.

Medinger, R., Nolte, V., Pandey, R. V., Jost, S., Ottenwälder, B., Schlötterer, C. & Boenigk, J. 2010. Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Mol. Ecol.*, 19:32–40.

Monier, A., Comte, J., Babin, M., Forest, A., Matsuoka, A. & Lovejoy, C. 2015. Oceanographic structure drives the assembly processes of microbial eukaryotic communities. *ISME J.*, 9:990–1002.

Montagnes, D. J. S., Roberts, E., Lukes, J. & Lowe, C. 2012. The rise of model protozoa. *Trends Microbiol.*, 20:184–191.

Moon-van der Staay, S. Y., De Wachter, R. & Vaulot, D. 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, 409:607–610.

Mordret, S., Romac, S., Henry, N., Colin, S., Carmichael, M., Berney, C., Audic, S., Richter, D. J., et al. 2015. The symbiotic life of *Symbiodinium* in the open ocean within a new species of calcifying ciliate (*Tiarina* sp.). *ISME J.*, 10:1424–1436.

Mori, A. S., Isbell, F. & Seidl, R. 2018. Beta-diversity, community assembly, and ecosystem functioning. *Trends Ecol. Evol.*, 33:549–564.

Nemergut, D. R., Schmidt, S., Fukami, T., O'Neill, S., Bilinski, T., Stanish, L., et al. 2013. Patterns and processes of microbial community assembly. *Microbiol. Mol. Biol. Rev.*, 77:342–356.

Nolte, V., Pandey, R. V., Jost, S., Medinger, R., Ottenwalder, B., Boenigk, J. & Schlötterer, C. 2010. Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Mol. Ecol.*, 19:2908–2915.

Oliverio, A. M., Power, J., Washburne, A., Cary, S., Stott, M. & Fierer, N. 2018. The ecology and diversity of microbial eukaryotes in geothermal springs. *ISME J.*, 12:1918–1928.

Orr, R. J., Zhao, S., Klaveness, D., Yabuki, A., Ikeda, K., Watanabe, M. & Shalchian-Tabrizi, K. 2018. Enigmatic Diphyllatea eukaryotes: culturing and targeted PacBio RS amplicon sequencing reveals a higher order taxonomic diversity and global distribution. *BMC Evol. Biol.*, 18:115.

Orsi, W., Charvet, S., Vdácny, P., Bernhard, J. M. & Edgcomb, V. 2012. Prevalence of partnerships between barcteria and ciliates in oxygen-depleted marine water columns. *Front. Microbiol.*, 3:341.

Orsi, W., Wilken, S., del Campo, J., Heger, T., James, E., Richards, T. A., Keeling, P. J., Worden, A. Z. & Santoro, A. E. 2018. Identifying protist consumers of photosynthetic picoeukaryotes in the surface ocean using stable isotope probing. *Environ. Microbiol.*, 20:815–827.

Parfrey, L. W., Walters, W. A., Lauber, C. L., Clemente, J. C., Berg-Lyons, D., Teiling, C., Kodira, C., Mohiuddin, M., Brunelle, J., Driscoll, M., Fierer, N., Gilbert, J. A. & Knight, R. 2014. Communities of microbial eukaryotes in the mammalian gut within the context of environmental eukaryotic diversity. *Front. Microbiol.*, 5:298.

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L. & Berney, C., et al. 2012. CBOL Protist working group: barcoding eukaryotic richness beyond animal, plant, and fungal kingdoms. *PLoS Biol.*, 10:e1001419.

Pedrós-Alió, C. 2006. Marine microbial diversity: can it be determined? *Trends Microbiol.*, 14:257–263.

Pedrós-Alió, C., Acinas, S., Logares, R. & Massana, R. 2018. Marine microbial diversity as seen by high throughput sequencing. In: Gasol, J. & Kirchman, D. (ed.), Microbial Ecology of the Oceans, 3rd ed. John Wiley & Sons, Hoboken. pp. 47–98.

Pernice, M. C., Giner, C. R., Logares, R., Perera-Bel, J., Acinas, S. G., Duarte, C. M., et al. 2016. Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J.*, 10:945–958.

Pitsch, G., Bruni, E. P., Forster, D., Qu, Z., Sonntag, B., Stoeck, T. & Posch, T. 2019. Seasonality of planktonic freshwater ciliates: are analyses based on v9 regions of the 18S rRNA gene correlated with morphospecies counts? *Front. Microbiol.*, 10. https://doi.org/10.3389/fmicb.2019.00248

Preston, F. W. 1948. The commonness, and rarity, of species. *Ecology*, 29:254–283.

de Queiroz, K. 2007. Species concepts and species delimitations. *Syst. Biol.*, 56:879–886.

Rengefors, K., Logares, R., Laybourn-Parry, J. & Gast, R. J. 2015. Evidence of concurrent local adaptation and high phenotypic plasticity in a polar microeukaryote. *Environ. Microbiol.*, 17:1510–1519.

Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ.*, 18:e2584.

Röttjers, L. & Faust, K. 2018. From hairballs to hypotheses—biological insights from microbial networks. *FEMS Microbiol. Rev.*, 42:761–780.

Salazar, G., Cornejo-Castillo, F. M., Benitez-Barrios, V., Fraile-Nuez, E., Alvarez-Salgado, X. A., Duarte, C. M., et al. 2016. Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME J.*, 10:596–608.

Salazar, G., Paoli, L., Alberti, A., Huerta-Cepas, J., Ruscheweyh, H.-J., Cuenca, M., Field, C. M., Coelho, L. P., Cruaud, C., Engelen, S., Gregory, A. C., Labadie, K., Marec, C., Pelletier, E., Royo-Llonch, M., Roux, S., Sánchez, P., Uehara, H., Zayed, A. A., Zeller, G., Carmichael, M., Dimier, C., Ferland, J., Kandels, S., Picheral, M., Pisarev, S., Poulain, J., Acinas, S. G., Babin, M., Bork, P., Boss, E., Bowler, C., Cochrane, G., de Vargas, C., Follows, M., Gorsky, G., Grimsley, N., Guidi, L., Hingamp, P., Iudicone, D., Jaillon, O., Kandels-Lewis, S., Karp-Boss, L., Karsenti, E., Not, F., Ogata, H., Pesant, S., Poulton, N., Raes, J., Sardet, C., Speich, S., Stemmann, L., Sullivan, M. B., Sunagawa, S. & Wincker, P. 2019. Gene expression hanges and community turnover differentially shape the global ocean metatranscriptome. *Cell*, 179:1068–1083.e21.

Santoferrara, L. F. 2019. Current practice in plankton metabarcoding: optimization and error management. *J. Plankton Res.*, 41:571–582.

Santoferrara, L. F., Grattepanche, J.-D., Katz, L. A. & McManus, G. B. 2014. Pyrosequencing for assessing diversity of eukaryotic microbes: analysis of data on marine planktonic ciliates and comparison with traditional methods. *Environ. Microbiol.*, 16:2752–2763.

Santoferrara, L. F., Grattepanche, J.-D., Katz, L. A. & McManus, G. B. 2016. Patterns and processes in microbial biogeography: do molecules and morphologies give the same answers? *ISME J.*, 10:1779–1790.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. & Weber, C. F. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, 75:7537–7541.

Ser-Giacomi, E., Zinger, L., Malviya, S., de Vargas, C., Karsenti, E., Bowler, C. & De Monte, S. 2018. Ubiquitous abundance distribution of non-dominant plankton across the global ocean. *Nat. Ecol. Evol.*, 2:1243–1249.

Shapiro, B. J. & Polz, M. F. 2014. Ordering microbial diversity into ecologically and genetically cohesive units. *Trends Microbiol.*, 22:235–247.

Sibbald, S. & Archibald, J. 2017. More protist genomes needed. *Nat. Ecol. Evol.*, 1:0145.

Stegen, J. C., Lin, X., Fredrickson, J. K., Chen, X., Kennedy, D. W., Murray, C. J., et al. 2013. Quantifying community assembly processes and identifying features that impose them. *ISME J.*, 7:2069–2079.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H. W. & Richards, T. A. 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.*, 19:21–31.

Stoeck, T., Behnke, A., Christen, R., Amaral-Zettler, L., Rodriguez-Mora, M., Chistoserdov, A., Orsi, W. & Edgcomb, V. 2009. Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC Biol.*, 7:72.

Stoeck, T., Breiner, H.-W., Filker, S., Ostermaier, V., Kammerlander, B. & Sonntag, B. 2014. A morpho-genetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. *Environ. Microbiol.*, 16:430–444.

Sunagawa, S., Coelho, L., Chaffron, S., Kultima, J., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., et al. 2015. Structure and function of the global ocean microbiome. *Science*, 348:1261359.

Taberlet, P., Bonin, A., Zinger, L. & Coissac, E. 2018: Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press, Oxford.

Tedersoo, L. & Anslan, S. 2019. Towards PacBio-based pan-eukaryote metabarcoding using full-length ITS sequences. *Environ. Microbiol. Rep.*, 11:659–668.

Tedersoo, L., Tooming-Klunderud, A. & Anslan, S. 2018. PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytol.*, 217:1370–1385.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., et al. 2015. Eukaryotic plankton diversity in the sunlit ocean. *Science*, 348:1261605.

Vasselon, V., Bouchez, A., Rimet, F., Jacquet, S., Trobajo, R., Corniquel, M., Tapolczai, K. & Domaizon, I. 2018. Avoiding quantification bias in metabarcoding: application of a cell biovolume correction factor in diatom molecular biomonitoring. *Meth. Ecol. Evol.*, 9:1060–1069.

Vellend, M. 2016: The Theory of Ecological Communities. Princeton University Press, Princeton.

Vermeij, G. J. & Dudely, R. 2000. Why are there so few evolutionary transitions between aquatic and terrestrial ecosystems? *Biol. J. Linn. Soc.*, 70:541–554.

Weber, A. A. T. & Pawlowski, J. 2013. Can abundance of protists be inferred from sequence data: a case study of foraminifera. *PLoS ONE*, 8:e56739.

Weinisch, L., Kirchner, I., Grimm, M., Kühner, S., Pierik, A., Roselló-Móra, R. & Filker, S. 2018b. Glycine betaine and ectoine are the major compatible solutes used by four different halophilic heterotrophic ciliates. *Microb. Ecol.*. https://doi.org/10.1007/s00248-018

Weinisch, L., Kühner, S., Roth, R., Grimm, M., Roth, T., Netz, D., Pierik, A. & Filker, S. 2018a. Identification of osmoadaptive strategies in the halophile, heterotrophic ciliate *Schmidingerothrix salinarum*. *PLoS Biol.*, 16:e2003892.

West, P. T., Probst, A. J., Grigoriev, I. V., Thomas, B. C. & Banfield, J. F. 2018. Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res.*, 28:569–580.

Wu, W., Lu, H. P., Sastri, A., Yeh, Y. C., Gong, G. C., Chou, W. C., et al. 2018. Contrasting the relative importance of species sorting and dispersal limitation in shaping marine bacterial versus protist communities. *ISME J.*, 12:485–494.

Yooseph, S., Nealsonm, K. H., Rusch, D. B., McCrow, J. P., Dupont, C. L., Kim, M., Johnson, J., Montgomery, R., Ferriera, S., Beeson, K., Williamson, S. J., Tovchigrechko, A., Allen, A. E., Zeigler, L. A., Sutton, G., Eisenstadt, E., Rogers, Y. H., Friedman, R., Frazier, M. & Venter, J. C. 2010. Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature*, 468:60–66.

Zhou, J. & Ning, D. 2017. Stochastic community assembly: does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.*, 81:e00002–e00017.