**Diversity and ecology of protists revealed by metabarcoding**

Fabien Burki1,2,*, Miguel M. Sandin1, and Mahwash Jamy1

1Department of Organismal Biology (Systematic Biology), Uppsala University, Norbyv. 18D, 75236 Uppsala, Sweden
2Science For Life Laboratory, Uppsala University, 75236 Uppsala, Sweden
*Correspondence: fabien.burki@ebc.uu.se
https://doi.org/10.1016/j.cub.2021.07.066

**SUMMARY**

Protists are the dominant eukaryotes in the biosphere where they play key functional roles. While protists have been studied for over a century, it is the high-throughput sequencing of molecular markers from environmental samples — the approach of metabarcoding — that has revealed just how diverse, and abundant, these small organisms are. Metabarcoding is now routine to survey environmental diversity, so data have rapidly accumulated from a multitude of environments and at different sampling scales. This mass of data has provided unprecedented opportunities to study the taxonomic and functional diversity of protists, and how this diversity is organised in space and time. Here, we use metabarcoding as a common thread to discuss the state of knowledge in protist diversity research, from technical considerations of the approach to important insights gained on diversity patterns and the processes that might have structured this diversity. In addition to these insights, we conclude that metabarcoding is on the verge of an exciting added dimension thanks to the maturation of high-throughput long-read sequencing, so that a robust eco-evolutionary framework of protist diversity is within reach.

**Introduction: Protists in a nutshell**

For all of the importance of understanding biodiversity, and how to preserve it in the face of rapid and global environmental changes, there is a category of organisms that is rarely discussed. These often small organisms — called protists — are much less studied than their macroscopic counterparts among eukaryotes (animals, plants, and fungi), yet they are increasingly recognised to play critical biogeochemical and ecological roles in most ecosystems. As microbes, protists represent the vast bulk of eukaryotic diversity, they live in virtually all environments on Earth, and their biomass at global scales is immense (estimated to be twice that of all animals)1–4. Understanding this diversity, and how it interacts with other microbes and larger organisms to form functioning ecosystems in all biomes, is fundamental to provide a holistic view of biological communities.

Before we move to the core concepts of this review, it is necessary to define what we actually mean by protists. This is because there is not a single accepted definition of what protists are and, as a fundamentally paraphyletic assemblage, protists represent a working concept rather than a biological entity5. Historically, protists have been widely regarded as a grab bag including anything eukaryote that is not an animal, land plant, or dikaryon fungus6. Although this is a definition by exclusion, it is useful because it clearly delineates — phylogenetically — what does or does not go in this grab bag. But it also means that the morphology, ecology, or more generally biology of protists encompass almost all of the broad spectrum of traits that we otherwise associate with eukaryotes. Indeed, protists are not only an essential component of today’s ecosystems, they have also been the only eukaryotic component for most of life’s history and as such have evolved into an incredible diversity of forms and functions1,7.

Protists are generally microscopic, unicellular eukaryotes but collectively they span more than six orders of magnitude in size (from eukaryotes smaller than bacteria to seaweeds taller than 10 story buildings). They can live solitary lives, preying on other small eukaryotes or prokaryotes, or form large colonies of a few meters. Protists can be purely phototrophic, in which case they are referred to as algae, or purely heterotrophic if they belong to the organisms traditionally called ‘protozoans’. In-between these two modes of nutrition, there are also a wide range of mixotrophic behaviours where phagotrophy and phototrophy co-exist in the same cells8. While some species can grow to very high densities if environmental conditions are favourable, most of the diversity remains regional and at low abundances; this diversity is known as the ‘rare-biosphere’9. Sometimes, blooms of microscopic protist algae are so widespread that they can be seen from space and, when taken collectively, these algae contribute as much to primary production as land plants10. More generally, protists are essential to ecosystem functions as partners in diverse symbioses (host and/or symbiont), as key regulating parasites (or parasitoids) of natural populations, and as intermediates in multiple trophic levels between small animals that prey on them and other microbes they consume11.

In recent years, high-throughput methods and large-scale sampling of different environments have provided unprecedented insights into the diversity and ecology of microbes. In this review, we discuss how the environmental DNA sequencing of protists has transformed our understanding of eukaryotic diversity as a whole, from revealing unprecedented lineage
Environmental DNA sequencing of protist diversity

Generally speaking, DNA (or RNA) sequencing can be done in two main ways depending on the source material: directed at individual taxa, for example from a culture or isolated organisms, or aimed at sequencing the entire DNA pool of a community — a process that is referred to as environmental sequencing. A community here represents many different entities within different environments and scales, from entire ecosystems (e.g., oceanic scales15, or a lake15, or soil in a forest or a field14,15) to host-associated microbiomes in a single specimen or a particular tissue16. Although environmental DNA (eDNA) is typically sampled with no other associated identifier than the environment itself, the sequencing of eDNA allows the detection and characterisation of natural communities without a priori knowledge of what members belong to these communities, in particular for analyzing protistan diversity and distributions in nature. As such, metabarcoding has resulted in a very large body of exciting insights from all environments, which form the core of this review. We will also focus on insights gained from the studies of the ‘total’ eukaryotic diversity of broadly defined ‘traditional’ environments such as soil, sediments, and water. The word ‘total’ here is meant to indicate the intention rather than the certain output of environmental sequencing, even when studies are designed to capture all diversity. Indeed, an array of methodological and biological reasons typically prevent the recovery of the full diversity present in an environment, or can recover the DNA of organisms not necessarily present or alive at the time of sampling (Box 2).

Metabarcoding

Figure 1. Environmental sequencing and metabarcoding.
Overview of environmental sequencing with emphasis on metabarcoding (see Box 1 for a description of the main steps). Environmental samples can be collected from any environment such as soil, sediment, water, air. Aquatic samples can be size filtered to study the different size fractions separately. Following total DNA and/or cDNA extraction, and depending on the aim of the study, metabarcoding and/or meta-genomics/meta-transcriptomics approaches can be carried out. For metabarcoding, eukaryotic primers (taxonomically broad or more restricted) are used to amplify targeted fragments (short or long) of the rDNA operon which are then sequenced. Sequence data are cleaned, clustered into operational taxonomic units (OTUs) or Amplicon Sequence Variants (ASVs) and used for taxonomic and ecological analyses.
Box 1. Steps in metabarcoding analysis of protist communities.

Metabarcoding (also known as amplicon sequencing, or sometimes metagenetics) consists of several interconnected steps that need to be critically assessed for correct interpretation. This box gives a brief overview of the main steps in a typical metabarcoding analysis, but extensive literature exists for more details. See Figure 1 for an illustration.

**Study design.** Arguably the most critical step when studying protist diversity is the scientific question. Depending on the scientific question asked, the study might for example focus on one or several environments, on a specific taxa or functional group, on a given time-frame or a temporal series. Consequently, the scientific question will determine the sampling strategy, i.e. the sampling location and frequency, the sampling method, the sample volumes, biological replicates, storing, and the bioinformatic analysis.

**Laboratory.** Metabarcoding requires the extraction of DNA (or RNA) from environmental samples. Commonly, DNA/RNA is extracted by breaking the cell, either chemically by detergents or physically by freezing and thawing. Thereafter, the targeted DNA region undergoes PCR amplification using sets of primers that might be more or less specific for particular taxonomic groups. Decisions on sequencing library preparation, reagents, multiplexing, or the choice of primers (Box 2) may all affect the amplification process. After amplification, the amplicons are sequenced on a high-throughput sequencing platform, the choice of which is typically based on a trade-off between three factors: the sequencing depth, the read length and the error rate.

**Bioinformatics.** After sequencing, the reads need thorough processing. This includes basecalling, removing low quality reads, pairing pair-end reads, demultiplexing, dereplication, de novo chimera identification, etc. Probably the most debated procedure of the read processing is read clustering or denoising into so-called Operational Taxonomic Units (OTUs) or Amplicon Sequence Variant (ASVs), respectively (Box 2). To avoid clustering, it is possible to work with raw reads after applying stringent quality and abundance filters. Extracting biologically meaningful reads is normally achieved by filtering based on arbitrary thresholds of abundance and/or presence on different samples. This step attempts to remove further artifacts and errors not identified in previous steps. Clean reads are also frequently assigned with taxonomy by comparison to reference databases, ideally curated (Box 2).

Later, it is possible to conduct the data analysis in order to explore biological patterns such as multivariate analysis, beta/gamma diversity, sequence similarity networks, etc.

**Replicability.** The last crucial step in metabarcoding analysis is the data deposition in public repositories or libraries. The data deposited should contain, at the very least, everything needed for the reproduction of the given study, from associated metadata to sequencing platform and technology. This step ensures that the analysis performed and the conclusions stated in the given study can be replicated, confirmed with different approaches or improved with new methods/algorithms.
Box 2. Limitations of metabarcoding, and solutions to deal with them.

Metabarcoding study design, execution, and analysis come with a range of potential technical and biological limitations. These limitations have been identified and can be addressed to unravel the full potential of metabarcoding. The most common source of biological limitation comes from the multicopy nature of the rDNA operon and the variability in the number of copies among different taxa. Copy number of the rDNA has been positively correlated to cell size. Therefore, derivation of cell abundance based on read counts should be interpreted with caution, especially for larger size fractions. For this reason (among others), metabarcoding analysis provides only relative information, or semiquantitative information after correction, which requires prior knowledge of the studied taxa. The multicopy nature of the rDNA operon not only affects abundance estimates, but also tends to inflate diversity estimates because of intragenomic polymorphism among the copies. Errors in DNA amplification or sequencing, tag-jumping during library preparation, cross-contamination during sequencing, or extracellular DNA may also contribute to overestimate the diversity. A recently developed tool, LULU, attempts to remove biologically redundant sequences (such as intra-genomic variability) by post-clustering reads based on similarity, abundances and co-occurrence. In order to lessen the impact of these errors, as well as to reduce redundant diversity and decrease computational effort, reads are traditionally clustered based on a similarity threshold, often 97–99% for protists, into OTUs. Several approaches exist to assemble OTUs. One popular approach is to use single nucleotide differences and read abundances in order to cluster based on natural thresholds (i.e., swarm). Other approaches are based on the direct estimate and correction of the expected error rate (denoising) of the given dataset, such as the Amplicon Sequence Variants (ASVs) produced by the DADA2 pipeline. At the same time as risks exist to inflate diversity estimates, the opposite problem — diversity underestimation — can also happen. Environmental surveys typically lack technical replicates (i.e., multiple sequencing runs of the same sample), which means that it is difficult to assess the true biological meaning of rare reads. Even though metabarcoding has become extremely high-throughput, undetected taxa are not necessarily proof of absence because truly low abundant taxa may be lost (removed through physical filtering), damaged (during storing) or not sequenced (stochastic factors). In addition, there are no truly universal primers covering all eukaryotic diversity. The presence of introns or insertions in the genetic marker can also hamper the sequencing of specific diversity. The selection of the primers (and therefore of the genetic marker) will affect the taxa that are successfully amplified, regardless of the natural abundance.

Lastly, reads are normally compared to a reference sequence (morphologically identified and annotated sequence) by either global or local alignment, resulting in a similarity score and/or a confidence estimate of the assignment. However, this step strongly relies on a comprehensive and carefully curated reference database, such as PR2 or SILVA, of sequences previously sequenced and annotated. The incompleteness of the reference database may lead to erroneous identifications of taxa. Therefore, the taxonomic assignment of environmental diversity has to be taken with caution. Other methods implement phylogeny-aware annotations in order to account for uncertainty and missing data in the reference database.

long-read technologies such as Pacific Bioscience (PacBio) or Oxford Nanopore Technologies (ONT). These technologies have recently produced high-quality long-read metabarcoding datasets of between 1,500 and 5,000 bp, although in theory the size limit is much higher than this range. Such datasets are appealing because they can easily contain the entire 18S rDNA gene, or even the full-length of the rDNA operon. This means that it is possible, in a single molecule, to cover many classical protist barcodes at once (e.g., V4, V9), as well as other markers in the 28S rDNA or the highly variable ITS region (Figure 2), while allowing for direct comparison to the vast body of knowledge provided by the 18S rDNA gene. The long metabarcodes can also be used in addition to reference sequences to produce denser and more robust reference phylogenies for the environmental diversity, allowing for direct phylogenetic comparison of the environmental diversity themselves. These are obvious advantages, yet it remains critical to fully assess the strengths versus weaknesses of long-read metabarcoding because with enhanced sequence length comes potential new biases. Among other issues, long insertions in some protist groups may prevent amplification, and there is potentially a higher risk of forming chimeric constructs for which specific detection tools need to be developed. Still, we have come a long way from the early days of low-accuracy long-read sequencing technologies, as today millions of long metabarcodes can be routinely produced with a quality similar to that of Sanger sequencing. In the future, it is likely that environmental sequencing will greatly benefit from the improved phylogenetic information contained in long metabarcoding datasets, especially in hybrid approaches that combine the current higher throughput and cheaper cost of short read metabarcoding.

Diversity and abundance of protists derived from environmental sequencing

Prior to the advent of environmental DNA sequencing, more than a century of protistology studies based on the isolation of organisms provided us with a rich account of what many of the main groups of protists are. This seminal and extensive body of work was first based on the morphology and functional characteristics of the isolated organisms (e.g., algae, small heterotrophs, absorptive nutrition), but was quickly complemented by molecular phylogenies when the sequencing of DNA became practicable. Nowadays, the study of the evolutionary relationships among the main groups of eukaryotes at the deepest phylogenetic levels almost always involves multiple genes, often hundreds of genes in a practice known as phylogenomics. These deep phylogenetic relationships have considerably changed over the decades, and continue to be challenged especially by the continuous flow of new genomic data for isolated
species but also increasingly from uncultured taxa; a description of the major eukaryotic groups, and their evolving relationships, have been recently discussed elsewhere. Remarkably, however, it is worth noting that the catalogue of these major protist groups has not been fundamentally altered since environmental sequencing, even high-throughput metabarcoding, came into play. In fact, less than a handful of high-ranking environmental lineages branching outside of established major groups have been discovered since the beginning of environmental sequencing (e.g., Picozoa, Rappemonads, or a novel group related to animals called MASHOL), some of which have now been assimilated to known groups (e.g., Rappemonads belong to haptophytes).

We can identify a few reasons as to why, perhaps against our initial expectations, high-throughput metabarcoding has not yielded the discovery of independent major groups of eukaryotes. After all, this is a method which by its sheer sequencing power could in theory sequence up to the last cells in any environment (but see Box 2). Firstly, as said above, the long history of microscopic descriptions and molecular phylogenetics of isolated taxa did a really good job at establishing the overall picture of protist diversity, so the emergence of metabarcoding built on an already richly populated diversity framework especially for the most abundant, economically important (e.g., pathogens), or easily identifiable (e.g., large cells, algae) taxa. But this rich historical account only partially explains the absence of novel high-ranking groups in metabarcoding datasets, because we know that such undescribed diversity exists from regular discoveries based on much lower-throughput, classical culturing approaches. For example, in the last five years, at least four major new branches in the tree of eukaryotes have been discovered or re-described, all from cultivation (Rigilids, Ancoracysta, Hemimastigophora, Rhodophidids).

We argue that the contrast between metabarcoding and cultivation in our ability to identify novel high-ranked diversity primarily stems from the current lack of phylogenetic signal in short read metabarcodes, and from the lack of reference sequences for groups that have not been discovered yet (logically). Indeed, most metabarcoding datasets do contain a proportion, often significant (e.g., about 30% in the oceans), of sequence diversity that cannot be reliably assigned to known groups due to high dissimilarity to references. This unassignable diversity mainly corresponds to the smallest and most rare protists, which coincidently are the organisms that have historically been the least studied. It is therefore reasonable to assume that among this unassignable diversity lie novel major eukaryotic groups awaiting characterization. Furthermore, there are other issues that prevent environmental sequences from being generated from taxa even if they exist. For example, some lifestyles are more likely to be underrepresented in environmental DNA surveys, notably obligate parasites of animals or endosymbionts in general, simply because those organisms are filtered out from common sampling practices that remove the larger size fractions corresponding to the hosts. So unless specifically targeted, host-associated organisms are often missed. Other inherent biases of metabarcoding, most importantly the choice of primers or the presence of insertions (Box 2), will select against species that are too divergent, resulting in some diversity remaining hidden unless specifically targeted. Lastly, while some environments have been heavily sampled (most notably the surface oceans), other environments remain severely under-explored and thus constitute an untapped reservoir of potential new diversity. These poorly studied environments not only correspond to far and difficult places to reach, such as the deep-sea or dense tropical forests, but also environments closer to us like lakes, ponds, rivers and soil, as well as whole hosts or specific tissues, that should all be the target of future sampling effort if we are to more fully grasp protist diversity.

So if not the discovery of novel high-ranked diversity, and amid the issues mentioned above, what have we learned about the diversity of protists from nearly 15 years of metabarcoding? The short answer is: a great deal, obviously. First and foremost, it is the expansion, generally massive, of the molecular diversity within already established groups that jumps to mind (Figure 3). Virtually every known group of protists saw its diversity soar with the addition of environmental lineages we had little suspicion...
The diversity and abundance of marine protists is thus known to harbor the richest protist communities on Earth, with rarefaction curves showing no sign of plateauing. Soil ecosystems are not only diverse, they also contain more unknown diversity than freshwater, and especially marine environments. Interestingly, comparisons across these ecosystems for the major protist groups have started to reveal extensive composition shifts, with terrestrial communities being much more similar to each other than to marine communities in terms of both diversity and abundance. As a result, some of the richest and most abundant groups in the oceans (e.g., radiolarians, diplonemids) are almost absent from terrestrial environments, while rare and low-diversity marine groups (most notably cercozoans and some amoebozoans) appear dominant in soils.

While most studies have thus far focused on the most abundant taxa in the environment, rare taxa (i.e., those with low local relative abundance) have also taken on a new significance as a result of metabarcoding. After a rocky start trying to weed out noise in the data (e.g., sequencing errors) and DNA present in the environment but not representative of active cells (e.g., moribund or dead cells, resting stages, or extracellular DNA), studies have shown that a metabolically active but rare protist biosphere is a common occurrence. In fact, this rare biosphere seems stable across communities, thus possibly representing an equilibrium shaped by biotic and abiotic interactions (although the ecological functions of the rare biosphere remain hypothetical). Strikingly, this rare biosphere contains even more diversity than...
more frequently encountered taxa, suggesting that abundant OTUs are always a minority in nature, while rare OTUs encompass the highest protist richness, up to a staggering 99% of all reported OTUs in some cases. Moreover, the majority of these rare OTUs seems to be evolutionarily related mainly to other rare OTUs, suggesting that some groups at least represent permanently rare taxa (whereas other taxa might increase in relative abundance in response to environmental changes). Because these rare taxa are among the least studied protists, and almost none have been characterised phylogenetically, it is likely that the rare biosphere includes the greatest potential for new lineage discoveries.

Taking all these results together, it is evident that the unprecedented breadth of metabarcoding data for protists has transformed our understanding of eukaryotic diversity in general, removing any possible doubts that these tiny organisms are the most diverse and abundant eukaryotes in the environment.

Global and local patterns of protist diversity, and the processes governing them

The breadth of metabarcoding data has not only given a better sense of the vast taxonomic diversity of protists, it has also provided new opportunities to effectively study how this diversity is structured within and across environments. This is a vital task to better understand the factors governing the assembly of protist communities, but also to predict how these communities will respond to global climate change. As often, a background of knowledge is available for animals and plants, but protist communities are structured following different patterns (described below; Figure 4). While the
dominant processes structuring these microbial communities are still largely unknown, a range of mechanisms has been proposed, including: biogeography driven by abiotic factors (e.g., salinity, pH, temperature, etc.) and biotic factors (interactions with other protist and non-protist members of the community), which can strongly influence which species are found in a habitat by filtering out ill-adapted ones; new species can be introduced into communities through dispersal, as well as speciation; and community composition can also change simply due to stochasticity (or sometimes referred to as ecological drift)\(^7\). These processes all act in conjunction but the relative importance of each varies depending on the scale of sampling and the environment. For example, protist communities separated by one meter will be less (or not at all) influenced by dispersal limitation than communities separated by hundreds of kilometers, and this effect will be less pronounced in marine environments than in more heterogeneous soil systems\(^8\). Below we briefly describe some of the main findings from protist ecology studies, namely endemicity of protist taxa, structuring of protist communities at global and local scales, and global patterns of species richness.

One of the main long-standing questions related to the organisation of microbial communities is how much of microbial diversity is cosmopolitan. A popular idea in the 1990s and early 2000s was formulated in the Baas Becking hypothesis, which stated that “Everything is everywhere, but the environment selects”\(^9\). This hypothesis proposed that the same microbial taxa will be found anywhere a suitable habitat exists, with no geographical limits, because microbes can disperse with no limitation due to their small size and high abundance. The availability of global-scale metabarcoding studies provided excellent opportunities to put this hypothesis to test, and it turned out that the vast majority of protists are in fact not cosmopolitan but instead have restricted geographical distributions (Figure 4), in part due to dispersal limitation\(^1\). For example, in the surface ocean only 0.35% of all OTUs (381 out of ∼110,000 OTUs) were deemed cosmopolitan\(^1\), and this was even less the case in soil (0.15% of cosmopolitan taxa)\(^2\). But what these cosmopolitan taxa lack in diversity, they make up for in abundance: these small protists (<20 \(\mu \text{m}\)) display biogeographic patterns at finer scales that are consistent with more limited Longhurst biogeographical provinces, each of which has a unique set of ecological conditions\(^3\). These biogeographic patterns thus suggest that smaller marine protists (especially phototrophs, for which these patterns are more pronounced) are more intimately linked with abiotic conditions such as nutrient availability and temperature than larger protists\(^4,5\). In contrast, soil and freshwater environments represent more discrete habitats than marine systems with abiotic factors showing a patchy distribution even at small scales. As a result, protist communities in these environments are more heterogeneous than in marine systems and are overall less predicted by geographic distance\(^6,7\). At a global scale, soil protist communities are better explained by environmental conditions, in particular mean annual precipitation, but other abiotic factors such as pH, temperature, and to some extent geographic distance also shape communities\(^8,9\). Co-occurrence patterns between protist and bacterial species have also been observed in soils and marine systems globally, highlighting the general importance of biotic interactions\(^10\).

At smaller spatial scales, both dispersal limitation and environmental conditions (such as temperature and salinity) contribute to protist community assembly in marine waters\(^11,12\). Thus, local oceanic features such as hydrothermal vents\(^13\) and anticyclonic gyres\(^14\) can lead to shifts in the communities even over short distances. For terrestrial ecosystems, the effects of dispersal limitation, which is typically not detected globally, become more manifest locally, for example in soils from different Neotropical forests\(^15\), lakes across Europe\(^16\), and paddy fields in East Asia\(^17\). However, the relative impact of dispersal limitation is variable, perhaps depending on the scale of the study, and environmental factors such as nutrient content, temperature, elevation, bacterial community, etc. also contribute to protist community assembly\(^18,11\). Furthermore, dispersal limitation and selection are not the only factors influencing protist communities; for instance, the protist communities in lakes in Eastern Antarctica seemed to have been assembled primarily through stochasticity\(^19\).

Finally, metabarcoding studies have also investigated patterns of species richness\(^20\). For instance, a large study of a marine macromictic latitudinal gradient from tropic to poles revealed a trend of diminishing diversity towards the pole, paralleling the poleward decline of local diversity known for larger organisms\(^21\) (Figure 4). The causes of this pattern are unclear, but it could be that protists in general are less tolerant of colder temperatures, and/or the warm tropics might be the sites of faster speciation rates due to higher kinetic energy and thereby faster mutation rates\(^22,23\). A similar pattern was not detected in the deeper ocean layer, probably because of a lower range of temperature differences between the poles and equator and increased nutrient availability\(^24\), nor has it been found in soils\(^25\). The elevational diversity gradient has been less studied, but there is some mixed evidence that protist species richness decreases with increasing elevation, as is the case for plants and animals\(^26,27\).

**Functional diversity of protists inferred from metabarcoding**

Beyond cataloguing protist diversity, and explaining organisational patterns of this diversity, one burning question raised by metabarcoding studies is: what does all this diversity actually do in the environment? Simply put, protists play the same broad...
functional roles as the rest of cellular life; that is they produce, degrade, and consume. While this is nothing new, the actual functional diversity of protists remains in many ways underappreciated \cite{1,105}. In reality, the enormous size range and taxonomic diversity of protists, their frequent abundance, their many nutritional modes, their complex morphology and behaviours, and their rapid metabolic rates result in key and complex functional roles in the first several trophic links of food webs. These roles have been best studied in marine systems \cite{9,11,105}, but they are also increasingly scrutinized in terrestrial environments \cite{10}.

The importance of primary production in the oceans for global carbon fixation, with marine unicellular algae accounting for about as much as terrestrial plants, is well established \cite{10}. More surprisingly, however, is an appreciation coming from comparisons of metabarcoding datasets across environments showing that the relative abundance of phototrophy in soil is not much lower than for marine plankton \cite{68}. If confirmed, this unexpectedly high proportion of soil phototrophy indicates a remarkable role of microbial photosynthesis outside of marine systems. Fixed carbon from primary production, in turn, becomes available for consumer herbivorous protists and small animals at several entry points in the food chains due to a vast array of pico- to micro-sized algal diversity \cite{3}. These heterotrophic consumers take a range of nutritional strategies, the best-known being predation by exclusive phagotrophs but mixotrophy (which are also producers, thus shifting the boundaries between trophic modes), parasitism, osmotrophy, and saprotrophy are all crucially interacting to structure ecosystems \cite{105}. In fact, heterotrophy is the over-dominant nutritional mode of protists \cite{12,29,68,109}, these organisms greatly contributing to control the population dynamics and community assembly of bacterial and eukaryotic preys alike \cite{107,108}. Among these heterotrophic protists, phagotrophic consumers account for at least half of all sequences in the ocean and on land \cite{12,29,68,69}. Parasites are also consistently to be extremely abundant (Figure 3): ballpark estimates show that in marine and soil samples, parasites represent up to 20% of all metabarcoding reads, supporting the idea that different parasite species likely infect every animal species on Earth \cite{14,69}. In marine environments, many of these parasites (or rather parasitoids, in which the host is always killed) belong to the syndiniales, a subgroup of the hyperdiverse MALVs, which are known to control blooms of dinoflagellates but can also infect other protists (e.g., ciliates, radiolarians), as well as animals \cite{109,110}. Land plants and fungi are also hosts of scores of parasites, as are other protists. In soil, the gregarine parasites of invertebrates (member of apicomplexans, which include for example the malaria agent \textit{Plasmodium}) were shown to be the most abundant OTUs in neotropical forests \cite{14}, and similar observations will probably be made everywhere else we look.

Protistan producers and consumers are thus not only taxonomically highly diverse and from all places in the eukaryotic tree, but they also represent a great functional diversity that is essential to ecosystem services. Deciphering the complexity of behaviours and lifestyles of these protists, who they are, and how exactly they interact not only among themselves but also with prokaryotes and larger organisms will be key to deepening our understanding of the functional roles of microbial eukaryotes.

**An eco-evolutionary approach to understand protist diversity**

The great taxonomic and functional diversity of protists outlined above is a result of hundreds of millions of years of evolution \cite{12,112}, perhaps more than two billion years since the origin of eukaryotes according to a recent estimate based on molecular clocks \cite{12}. Therefore, to study how this diversity is generated and maintained not only requires looking at the current patterns of diversity, but also turning to the past to look at diversification processes throughout geological times. This diversification — the balance between speciation and lineage extinction — is, however, difficult to study because direct organismal observations are typically lacking for most of life’s history. Fossils have been used for many decades as a direct means to estimate diversification dynamics in a multitude of eukaryotic groups, including protist groups using their continuous and well-documented microfossil records \cite{113-115}. But while we have gained some major insights from these palaeontological studies \cite{116,117}, the vast majority of eukaryote diversity is non-fossilized and existing fossils can be hard to interpret, particularly those very ancient microfossils from the Proterozoic Era (ca. 2000–1400 to 720 Ma) \cite{118,119}.

However, phylogenetic comparative methods implementing stochastic models of diversification based on extant diversity have flourished in recent years \cite{120}. These methods use lineage relationships and contemporary trait information to look at how these traits have evolved and which factors (e.g., paleoenvironmental) have influenced diversification \cite{121}. No method is perfect, however, and purely phylogenetic-based studies of diversification have raised concerns that the extrapolation of extinction rates from only living taxa cannot be reliably made \cite{122,123}.

As often, it is a combination of approaches and data that have helped to gain a deeper understanding of how ecological and evolutionary processes have generated the biodiversity we observe today. Indeed, joint interpretations of both molecular data from extant specimens and the palaeontological record have shown increasingly coherent diversification histories and represent an exciting avenue of research, including the diversification of important fossilizable protist groups such as diatoms \cite{124,125}, dinoflagellates \cite{126-128}, foraminifera \cite{129}, radiolarians \cite{128,129}, and testate amoebae \cite{130}. The diversification of these ecologically dominant groups appears to have occurred relatively suddenly in the late Neoproterozoic (ca. 650 Ma) fossil record, but before this time early eukaryote evolution seems to have been much calmer. Eukaryotes during the Proterozoic showed a generally low standing diversity in unfavorable conditions while most of the biological activity was likely governed by prokaryotes \cite{131}. This period is in fact sometimes referred to as the ‘boring billion’ (ca. 1800–800 Ma) to reflect the rarity of eukaryotic fossils and the constantly low oxygen concentration \cite{132}.

It was, however, during this period of low eukaryote activity that the stage was set for the later expansion to hyperdominance of some protist groups, notably by the establishment of plastid endosymbioses that gave rise to a variety of algae (e.g., dinoflagellates, diatoms) that profoundly altered the global geochemical and ecological conditions of the Earth \cite{112,133}. More generally, species interactions of many types (i.e., symbiosis such as parasitism or predation) have likely played critical roles in the overall diversification of eukaryotes during the late Proterozoic \cite{134}. For example, it was suggested that predation from large unicellular...
rhizarians (e.g., Foraminifera and Radiolaria) contributed to important ecological changes, from the continuous proliferation of algae\textsuperscript{213} to perhaps an arm race with animals leading to the development of multicellularity in this group\textsuperscript{135}. Conversely, the predatory pressure from animals may have led to the development of extracellular skeletons in many protist groups\textsuperscript{136,137}. During the Jurassic (ca. 200–145 Ma), dinoflagellates established photosymbiosis with corals\textsuperscript{138}. In this period, Foraminifera and Radiolaria also developed endosymbiotic relationships with photosynthetic dinoflagellates and haptophytes likely in response to oligotrophic oceans, bringing these associations to the planktonic realm\textsuperscript{139,140}. On land, organismal interactions involving protists also critically contributed to the diversification processes. For example, it has been suggested that interactions between bacteria, plants and testate amoebae led to the development of a complex rhizosphere and the necessary biogeochemical changes for the expansion of land plants during the early Phanerozoic\textsuperscript{141}.

Metabarcoding data have much to contribute to these eco-evolutionary studies of protist diversification, because their throughput currently represents the best approximation of OTU-level diversity in many environments. This is important because a near-full sampling of extant diversity offers the best possibilities for fitting diversification models using phylogenetic approaches\textsuperscript{120}. The use of metabarcoding data in eco-evolutionary studies is in its infancy, but it has already shown great promise in spite of the limited phylogenetic signal in short-read metabarcodes. For example, combining phylogenetic models of diversification with palaeoenvironmental data, a recent study showed a strong effect of specific environmental conditions (variation in carbon dioxide partial pressure) on early diatom evolution\textsuperscript{124}. With the current development of long-read metabarcoding, more robust phylogenies including large proportions of the environmental diversity will become available, therefore allowing more detailed analyses of diversification dynamics across the diversity of eukaryotes.

**Concluding remarks**

The last 15 years of research on protist diversity have seen a great deal of inspiring insights. Our understanding of the vastness of this diversity, how and by which processes it is structured through space and time, and what its functional roles are, is changing so rapidly that it can be challenging to keep pace. Excitingly, there is no sign that this pace is slowing down. Metabarcoding has massively contributed to this new perspective on protist diversity, allowing deep sequencing of an always broader range of environments. While this approach will likely remain dominant in diversity studies, it will also evolve together with new sequencing technologies, as it has in the past. Notably, we anticipate that long-read sequencing will lead to a better integration of the different molecular markers commonly used, greatly facilitating the comparison of all the data we have produced and will continue to produce. Other methods will play key roles too, such as metagenomics that not only allows recovery of taxonomic markers without potential amplification biases, but also offers a powerful way to integrate more parameters (e.g., metabolism). Beyond molecules, however, it is important to emphasize that we actually have no idea what the vast majority of this protistan diversity looks like, and how it lives, from lack of direct observations. Therefore, a critical next step in the study of protist diversity will be to devise strategies to link this catalogue of environmental sequences to images and genomes, and ideally cultures, in order to increase the predictive power of metabarcoding data and in turn more fully grasp the ecological functions of protists.

**ACKNOWLEDGEMENTS**

F.B.’s research is supported by a Fellowship grant from SciLifeLab, a research project grant from the Swedish Research Council (VR-2017-04663), and a Future Research Leader grant from Formas (201701197). We would like to thank Christophe Seppey and the rest of the authors from the study ‘Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems’\textsuperscript{315} for kindly sharing their data and helpful scripts to produce Figure 3, and for being very responsive to our emails.

**DECLARATION OF INTERESTS**

The authors declare no competing interests.

**REFERENCES**

1. Keeling, P.J., and Burki, F. (2019). Progress towards the Tree of Eukaryotes. Curr. Biol. 29, R808–R817.
2. Geisen, S., Mitchell, E.A.D., Adl, S., Borkowski, M., Dunthorn, M., Ekelund, F., Fernández, L.D., Jousset, A., Krashevska, V., Singer, D., et al. (2018). Soil protists: a fertile frontier in soil biology research. FEMS Microbiol. Rev. 42, 293–323.
3. Caron, D.A., Countway, P.D., Jones, A.C., Kim, D.Y., and Schnetzer, A. (2012). Marine protistan diversity. Annu. Rev. Mar. Sci. 4, 467–493.
4. Bar-On, Y.M., Phillips, R., and Milo, R. (2018). The biomass distribution on Earth. Proc. Natl. Acad. Sci. USA 115, 6506–6511.
5. O’Malley, M.A., Simpson, A.G.B., and Roger, A.J. (2013). The other eukaryotes in light of evolutionary protistology. Biol. Philos. 28, 299–330.
6. Whittaker, R.H. (1969). New concepts of kingdoms of organisms. Science 163, 150–160.
7. Burki, F., Roger, A.J., Brown, M.W., and Simpson, A.G.B. (2020). The new tree of eukaryotes. Trends Ecol. Evol. 35, 43–55.
8. Selosse, M.-A., Charpin, M., and Not, F. (2017). Mixotrophy everywhere on land and in water: the grand écart hypothesis. Ecol. Lett. 20, 246–263.
9. Logares, R., Mangot, J.-F., and Massana, R. (2015). Rarity in aquatic micro: placing protists on the map. Res. Microbiol. 166, 831–841.
10. Falkowski, P. (2012). Ocean Science: The power of plankton. Nature 483, S17–S20.
11. 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., et al. (2017). Probing the evolution, ecology and physiology of marine protists using transcriptomics. Nat. Rev. Microbiol. 15, 8–20.
12. de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Bescot, N.L., Probert, I., et al. (2015). Eukaryotic plankton diversity in the sunlit ocean. Science 348, 1261605.
13. David, G.M., Moreira, D., Rebull, G., Annenkov, N.V., Galindo, L.J., Bertoño, P., López-Archilla, A.I., Jardillier, L., and López-García, P. (2021). Environmental drivers of plankton protist communities along latitudinal and vertical gradients in the oldest and deepest freshwater lake. Environ. Microbiol. 23, 1436–1451.
14. Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., Singer, D., Mayor, J., Bunge, J., Semaka, S., et al. (2017). Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. Nat. Ecol. Evol. 1, 1–8.
15. Luan, L., Jiang, Y., Cheng, M., Dini-Andreote, F., Sui, Y., Xu, Q., Geisen, S., and Sun, B. (2020). Organism body size structures the soil microbial and nematode community assembly at a continental and global scale. Nat. Commun. 11, 6406.

16. Sapp, M., Ploch, S., Fiore-Donno, A.M., Bonkowski, M., and Rose, L.E. (2018). Protests are an integral part of the Arabidopsis thaliana microbiome. Environ. Microbiol. 20, 30–43.

17. Siapeta, J., López-Garcia, P., and Moreira, D. (2006). Global dispersal and ancient cryptic species in the smallest marine eukaryotes. Mol. Biol. Evol. 23, 23–29.

18. De Schepper, S., Ray, J.L., Skaar, K.S., Sadatzki, H., Ijaz, U.Z., Stein, R., and Larsen, A. (2019). The potential of sedimentary ancient DNA for re-constructing past sea ice ecosystems. ISME J. 13, 2566–2577.

19. Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.-R., Labadie, K., Salazar, G., Dini-Andreote, F., Sui, Y., Xu, Q., Geisen, S., and Larsen, A. (2019). Unexpected diversity of small eukaryotes in deep-sea Antarctic and nematode community assembly at a continental and global scale. ISME J. 9, 1045–1058.

20. Salazar, G., Paoli, L., Alberti, A., Huerta-Cepas, J., Ruscheweyh, H.-J., Pernice, M.C., Logares, R., Guillou, L., and Massana, R. (2013). Protists are an integral part of the Arabidopsis thaliana microbiome. Mol. Ecol. Resour. 13, 41–45.

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

22. Obioli, A., Giner, C.R., Sánchez, P., Duarte, C.M., Acinas, S.G., and Massana, R. (2020). A metagenomic assessment of microbial eukaryotic diversity in the global ocean. Mol. Ecol. Resour. 20, 718–731.

23. Balsalobre, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Lev esque, C.A., Chen, W., and Consortium, F.B. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc. Natl. Acad. Sci. USA 109, 8241–8246.

24. Reboleau, G.P., De Cock, A.W.A.M., Coopey, M.D., Voglmayr, H., Brouwer, H., Baia, K., Chitty, D.W., Desauniers, N., Eggertson, Q.A., Gachon, C.M.M., et al. (2011). DNA barcoding of oomycetes with cytochrome c oxidase subunit I and internal transcribed spacer. Mol. Ecol. Resour. 11, 1002–1011.

25. Newmaster, S.G.N.G., Fazekas, A.J.F.J., and Ragupathy, S.R. (2006). DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. Can. J. Bot. 84, 335–341.

26. Bailet, B., Bouchez, A., Franc, A., Frigerio, J.-M., Keck, F., Karjalainen, S.-M., Rimet, F., Schneider, S., and Kahler, M. (2019). Molecular versus morphological data for benthic diatoms biomonitoring in Northern European freshwater and consequences for ecological status. Metabarcoding and Metagenomics 3, e34002.

27. Gall, L.L., and Saunders, G.W. (2010). DNA barcoding is a powerful tool to uncover algal diversity: A case study of the Phyllophoraceae (gigartinales, Rhodophyta) in the Canadian Flora 1. J. Phycol. 46, 374–389.

28. Kucera, H., and Saunders, G.W. (2008). Assigning morphological vari-ants of Fucus (Fucales, Phaeophyceae) in Canadian waters to recogni-zed species using DNA barcoding. Botany 86, 1065–1079.

29. Balsalobre, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Lev esque, C.A., Chen, W., and Consortium, F.B. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc. Natl. Acad. Sci. USA 109, 8241–8246.

30. Kronk, P.D., Cuvyskins, A., Ball, S.L., and deWard, J.R. (2003). Biological identifications through DNA barcodes. Proc. Biol. Sci. 270, 313–321.

31. Hillis, D.M., and Dixon, M.T. (1991). Ribosomal DNA: Molecular evolution and phylogenetic inference. Q. Rev. Biol. 66, 411–433.

32. López-Garcia, P., Rodriguez-Valera, F., Pedró-Alí, C., and Moreira, D. (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature 409, 603–607.

33. Moon-van der Staa, S.Y., De Wachter, R., and Vaulot, D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409, 607–610.

34. Not, F., Gaulsing, R., Azam, F., Heidelberg, J.F., and Worden, A.Z. (2007). Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. Environ. Microbiol. 9, 1233–1252.

35. Massana, R., Castresana, J., Balaguer, V., Guillou, L., Romari, K., Grosil lieur, A., Valentin, K., and Pedrós-Alió, C. (2004). Phylogenetic and ecolog-ical analysis of novel marine Stramenopiles. Appl. Environ. Microbiol. 70, 3528–3534.

36. Pernice, M.C., Logares, R., Guillou, L., and Massana, R. (2013). General patterns of diversity in major marine microeukaryote lineages. PLoS One 8, e57170.

37. Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., and Huse, S.M. (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.

38. Pernice, M.C., Giner, C.R., Logares, R., Perera-Bel, J., Acinas, S.G., Duarte, C.M., Gasol, J.M., and Massana, R. (2016). Large variability of bathypelagic microbial eukaryotic communities across the world’s oceans. ISME J. 10, 945–958.

39. Oliverio, A.M., Geisen, S., Delgado-Baquerizo, M., Maestre, F.T., Turner, B.L., and Fierer, N. (2020). The global-scale distributions of soil protists and their contributions to belowground systems. Sci. Adv. 6, eaax8778.

40. Boenigk, J., Wodniok, S., Bock, C., Hempel, C., Grossmann, L., Lange, A., and Jensen, M. (2018). Geographic distance and mountain ranges structure freshwater protist communities on a European scale. MBMG Metabarcoding Metagenomics, 3269–3266.

41. Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., et al. (2012). The Protist Ribo-soomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 41, D597–D604.

42. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2012). The SILVA ribosomal RNA gene data-base project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596.

43. Berger, S.A., Krompass, D., and Stamatakis, A. (2011). Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. Syst. Biol. 60, 291–302.

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

45. Duarte, C.M., Gasol, J.M., and Massana, R. (2016). Large variability of eukaryotic diversity and distribution. PLoS One 11, e0150025.

46. Örjan, A., Houten, S., Klaveness, D., Yabuki, A., Ikeda, K., Watanabe, M., and Balch, W.B. (2010). Improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596.

47. Tedersoo, L., Tooming-Klaerner, A., and Anslan, S. (2018). PacBio metabarcoding of Fusig and other eukaryotes: errors, biases and perspectives. New PhytoL 217, 1370–1385.
Longhurst, A. (2006). Ecological Geography of the Sea (Amsterdam: Elsevier).

Pinseel, E., Janssens, S.B., Verleyen, E., Vanormelingen, P., Kohler, T.J., Biersma, E.M., Sabbe, K., Van de Vijver, B., and Vyverman, W. (2020). Global radiation in a rare biosphere soil diatom. Nat. Commun. 11, 2382.

Lima-Mendes, G., Faust, K., Henry, N., Collin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J.C., Roux, S., Vincent, F., et al. (2015). Determinants of community structure in the global plankton interactome. Science 348, 1262073.

Allen, R., Summerfield, T., Currie, K., Dillingham, P., and Hoffmann, L. (2020). Distinct processes structure bacterioplankton and protist communities across an oceanic front. Aquat. Microb. Ecol. 85, 19–34.

Wu, W., Lu, H.-P., Sastri, A., Yeh, Y.-C., Gong, G.-C., Chou, W.-C., and Chaffron, S., Ignacio-Espinosa, J.C., Roux, S., Vincent, F., et al. (2015). Determinants of community structure in the global plankton interactome. Science 348, 1262073.

Worden, A.Z., Follmann, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E., and Keeling, P.J. (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. Science 347, 1257594.

Seeleuthner, Y., Mondy, S., Lombard, V., Carradec, Q., Pelletier, E., Wessmer, M., Lecomte, J., Mangot, J.-F., Poulain, J., Labadie, K., et al. (2018). Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. Nat. Commun. 9, 310.

Hüninghaus, M., Koller, R., Kramer, S., Marhan, S., Kandelé, E., and Bonkowski, M. (2017). Changes in bacterial community composition and soil respiration indicate rapid successions of protist grazers during mineralization of maize crop residues. Pedobiologia 62, 1–8.

Geisen, S., Mitchell, E.A.D., Wilkinson, D.M., Adl, S., Bonkowski, M., Brown, M.W., Fiore-Donno, A.M., Heger, T.J., Jassey, V.E.D., Krasheska, V., et al. (2017). Soil protistology rebooted: 30 fundamental questions to start with. Soil Biol. Biochem. 111, 94–103.

Guillou, L., Viprey, M., Chambouvet, A., Walsh, R.M., Kirkham, A.R., Massara, R., Scanlan, D.J., and Worden, A.Z. (2008). Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndicales (Alveolata). Environ. Microbiol. 10, 3349–3365.

Chambouvet, A., Morin, P., Marie, D., and Guillou, L. (2008). Control of toxic marine dinoflagellate blooms by serial parasitic killers. Science 322, 1254–1257.

Knoll, A. H., Javaux, E. J., Hewitt, D., and Cohen, P. (2006). Eukaryotic organisms in Proterozoic oceans. Philos. Trans. R. Soc. B Biol. Sci. 361, 1023–1038.

Strassert, J.F.H., Irissari, I., Williams, T.A., and Burki, F. (2021). A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. Nat. Commun. 12, 1879.

Bown, P.R., Lees, J.A., and Young, J.R. (2004). Calcareous nanoplankton evolution and diversity through time. In Coccolithophores: From Molecular Processes to Global Impact, H.R. Thierstein, and J.R. Young, eds. (Springer), pp. 481–508.

Rabosky, D.L., and Sorhannus, U. (2009). Diversity dynamics of marine planktonic diatoms across the Cenozoic. Nature 457, 183–186.

Lazarus, D., Barron, J., Renaudie, J., Diver, P., and Türke, A. (2014). Cenozoic planktonic marine diatom diversity and correlation to climate change. PloS One 9, e84857.

Ezard, T.H.G., Aze, T., Pearson, P.N., and Purvis, A. (2011). Interplay between changing climate and species’ ecology drives macroevolutionary dynamics. Science 332, 349–351.

Yasuhara, M., Tittensor, D.P., Hillebrand, H., and Worm, B. (2017). Combining marine macroecology and palaeoecology in understanding biodiversity: microfossils as a model. Biol. Rev. 92, 199–215.

Eme, L., Sharpe, S.C., Brown, M.W., and Roger, A.J. (2014). On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. Cold Spring Harb. Perspect. Biol. 6, a016139.

Donoghue, P.C.J. (2020). Fossil cells. Curr. Biol. 30, R485–R490.

Morlon, H. (2014). Phylogenetic approaches for studying diversification. Ecol. Lett. 17, 508–525.

Cornwell, W., and Nakagawa, S. (2017). Phylogenetic comparative methods. Curr. Biol. 27, R333–R336.

Quental, T.B., and Marshall, C.R. (2011). The molecular phylogenetic signature of clades in decline. PLoS One 6, e25780.

Quental, T.B., and Marshall, C.R. (2010). Diversity dynamics: molecular phylogeny need the fossil record. Trends Ecol. Evol. 25, 434–441.

Lewitus, E., Bittrner, L., Malviya, S., Bowler, C., and Morlon, H. (2018). Clade-specific diversification dynamics of marine diatoms since the Jurassic. Nat. Ecol. Evol. 2, 1715–1723.

Nakov, T., Beaulieu, J.M., and Alversen, A.J. (2019). Diatoms diversity and turn over faster in freshwater than marine environments. Evolution 73, 2497–2511.
126. Janouškové, J., Gavelis, G.S., Burki, F., Dinh, D., Bachvaroff, T.R., Gornik, S.G., Bright, K.J., Imanian, B., Strom, S.L., Delwiche, C.F., et al. (2017). Major transitions in dinoflagellate evolution unveiled by phylogenomics. Mol. Phylogenet. Evol. 114, E171–E180.

127. Groussin, M., Pawlowski, J., and Yang, Z. (2011). Bayesian relaxed clock estimation of divergence times in foraminifera. Mol. Phylogenet. Evol. 67, 157–166.

128. Sandin, M.M., Pillet, L., Biard, T., Poirier, C., Bigeard, E., Romac, S., Suzuki, N., and Not, F. (2019). Time calibrated morpho-molecular classification of Nassellaria (Radiolaria). Protist 170, 187–208.

129. Sandin, M.M., Biard, T., Romac, S., O’Dogherty, L., Suzuki, N., and Not, F. (2021). A morpho-molecular perspective on the diversity and evolution of Spumellaria (Radiolaria). Protist 172, 125906.

130. Porter, S.M., and Knoll, A.H. (2000). Testate amoebae in the Neoproterozoic Era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. Paleobiology 26, 360–385.

131. Cohen, P.A., and Macdonald, F.A. (2015). The proterozoic record of eu-karyotes. Paleobiology 41, 610–632.

132. Holland, H.D. (2006). The oxygenation of the atmosphere and oceans. Philos. Trans. R. Soc. B Biol. Sci. 361, 903–915.

133. Mukherjee, I., Large, R.R., Corkrey, R., and Danyushevsky, L.V. (2018). The Boring Billion, a slingshot for complex life on Earth. Sci. Rep. 8, 4432.

134. Brocks, J.J., Jarrett, A.J.M., Sirantoine, E., Hallmann, C., Hoshino, Y., and Liyanage, T. (2017). The rise of algae in Cryogenian oceans and the emergence of animals. Nature 548, 578–581.

135. Nettersheim, B.J., Brocks, J.J., Schwelm, A., Hope, J.M., Not, F., Lomas, M., Schmidt, C., Schiebel, R., Nowack, E.C.M., De Deckker, P., et al. (2019). Putative sponge biomarkers in unicellular Rhizaria question an early rise of animals. Nat. Ecol. Evol. 3, 577–581.

136. S. Bengtson, ed. (1994). Early Life on Earth: Nobel Symposium, No. 84 (Columbia University Press).

137. Porter, S. (2011). The rise of predators. Geology 39, 607–608.

138. LaJeunesse, T.C., Parkinson, J.E., Gabrieleosn, P.W., Jeong, H.J., Reimer, J.D., Voolstra, C.R., and Santos, S.R. (2018). Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. Curr. Biol. 28, 2570–2580.e6.

139. Shaked, Y., and de Vargas, C. (2006). Pelagic photosymbiosis: rDNA assessment of diversity and evolution of dinoflagellate symbionts and planktonic foraminiferal hosts. Mar. Ecol. Prog. Ser. 325, 59–71.

140. Decelle, J., Probert, I., Bittner, L., Desveives, Y., Colin, S., Vargas, C., de Gali, M., Simó, R., and Not, F. (2012). An original mode of symbiosis in open ocean plankton. Proc. Natl. Acad. Sci. USA 109, 18000–18005.

141. Lahiri, J.J., Bosak, T., Lara, E., and Mitchell, E.A.D. (2015). The Phanerozoic diversification of silica-cycling testate amoebae and its possible links to changes in terrestrial ecosystems. PeerJ 3, e1234.

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

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

144. Harrison, J.B., Sunday, J.M., and Rogers, S.M. (2019). Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proc. R. Soc. B Biol. Sci. 286, 20191409.

145. Pittsch, G., Bruni, E.P., Forster, D., Qu, Z., Sonntag, B., Stoeck, T., and Posch, T. (2019). Seasonality of planktonic freshwater ciliates: are analyses based on V9 regions of the 18S rRNA gene correlated with morphology species counts? Front. Microbiol. 10, 248.

146. Caron, D.A., and Hu, S.K. (2019). Are we overestimating protistan diversity in nature? Trends Microbiol. 27, 197–205.

147. Schnell, I.B., Bohmann, K., and Gilbert, M.T.P. (2015). Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. Mol. Ecol. Res. 15, 1289–1303.

148. Fraslev, T.G., Kjell, R., Bruun, H.H., Eijmaas, R., Brunberg, A.K., Pietroni, C., and Hansen, A.J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nat. Commun. 8, 1186.

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

150. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583.

151. Vaulot, D., Geisen, S., Mahé, F., and Bass, D. (2021). pr2-primers: an 18S rRNA primer database for protists. Mol. Ecol. Resour. https://doi.org/10.1111/1755-0998.13465.

152. Bernier, C.R., Petrov, A.S., Kovacs, N.A., Penev, P.I., and Williams, L.D. (2018). Translation: the universal structural core of life. Mol. Biol. Evol. 35, 2065–2076.