Trends in Microbiology

CellPress REVIEWS

Opinion

Are We Overestimating Protistan Diversity in Nature?

David A. Caron[1,*] and Sarah K. Hu[1]

Documenting the immense diversity of single-celled, eukaryotic organisms (protists) has been a formidable challenge for ecologists. These species were originally defined by morphological criteria, but shortcomings of the morpho-species concept, and a bewildering array of sizes and cellular attributes, has made constructing a taxonomy that is useful for ecologists nearly impossible. Consequently, physiological and genetic information has been integrated to address these shortcomings, and to develop the framework of a unifying taxonomy. DNA sequence information, in particular, has revolutionized studies of protistan diversity. However, the exponential increase in sequence-based protistan species richness published from field surveys in recent years raises the question of whether we have moved beyond characterizing species-level diversity and begun to reveal intraspecies diversity. The answer to that question appears to be 'yes', at least for some protistan lineages. The need to document such microdiversity may be justified, but it is important for protistologists to recognize and acknowledge that possibility, and its consequences.

Highlights

Until recently it has been difficult to determine the full breadth of protistan species richness in natural communities using traditional morphology-based taxonomy, but high-throughput sequencing (HTS) of target genes has revealed unprecedented and unexpected species richness.

Paradoxically, some recent estimates of the number of protistan operational taxonomic units (OTUs) are now approaching the total number of individual protists present in small samples. Many protistologists appear to be unaware of this paradox in their quest to fully document the diversity of natural microbial communities.

Intraspecies variability can be an important component of the behavior and physiology of many protistan morphospecies.

A path forward is needed that differentiates interspecies from intraspecies sequence variability, including strict quality control on sequences and standardization of OTU-calling methods.

Protists, Their Diversity, and Why They Are Important

Single-celled, eukaryotic organisms (protists) are responsible for numerous essential ecological and biogeochemical activities in terrestrial and aquatic environments, including photosynthesis, trophic coupling, elemental transformations, decomposition, and disease. Photosynthetic protists form the base of food chains in important fisheries throughout the world [1,2], while heterotrophic forms are essential in the decomposition and remineralization of nutrient elements in water, sediments, and soils [3–5]. Characterizing protistan diversity is fundamentally important for understanding their biogeography, ecology, and biogeochemical significance in modern (and future) ecosystems. Additionally, the fossilized remains of many species (e.g., tests, scales, etc.) have been employed as tools for reconstructing climatological conditions of ancient oceans and lakes [6].

Protists constitute an incredible diversity of form and function within the eukaryotic tree of life that also contains the animals, plants, and fungi, albeit these latter groups form rather modest branches in comparison to their unicellular brethren [7,8]. They exist in nearly every ecosystem on Earth that harbors life, with many thousands of described species, including many fossil 'species'. While most protistologists believe that vast numbers of protistan species still await discovery and description, some suggest that most major 'types' of protists have already been documented. These differing views reveal a fundamental controversy in defining protistan species.

A Complicated History of Defining Protistan Species

Describing the diversity of protists in nature has been a difficult endeavor since their discovery approximately three and half centuries ago by Antonie van Leeuwenhoek [9], and it remains a

[1]Department of Biological Sciences, 3616 Trousdale Parkway, University of Southern California, Los Angeles, CA 90089-0371, USA

*Correspondence: dcaron@usc.edu (D.A. Caron).

daunting challenge today. One recent study estimated species richness on our planet at more than one trillion species [10] with most of these, of course, being microorganismal species including an unknown but significant proportion of protists. The basis of these estimates depends on equating 'genetic' (i.e., DNA sequence-based) diversity with 'species' diversity. Such interpretations have become common for bacteria and archaea [11], but only recently have they been widely applied to protists. The slower adoption of a sequence-based species concept exists, in part, because protistan species have traditionally been defined based on morphological features.

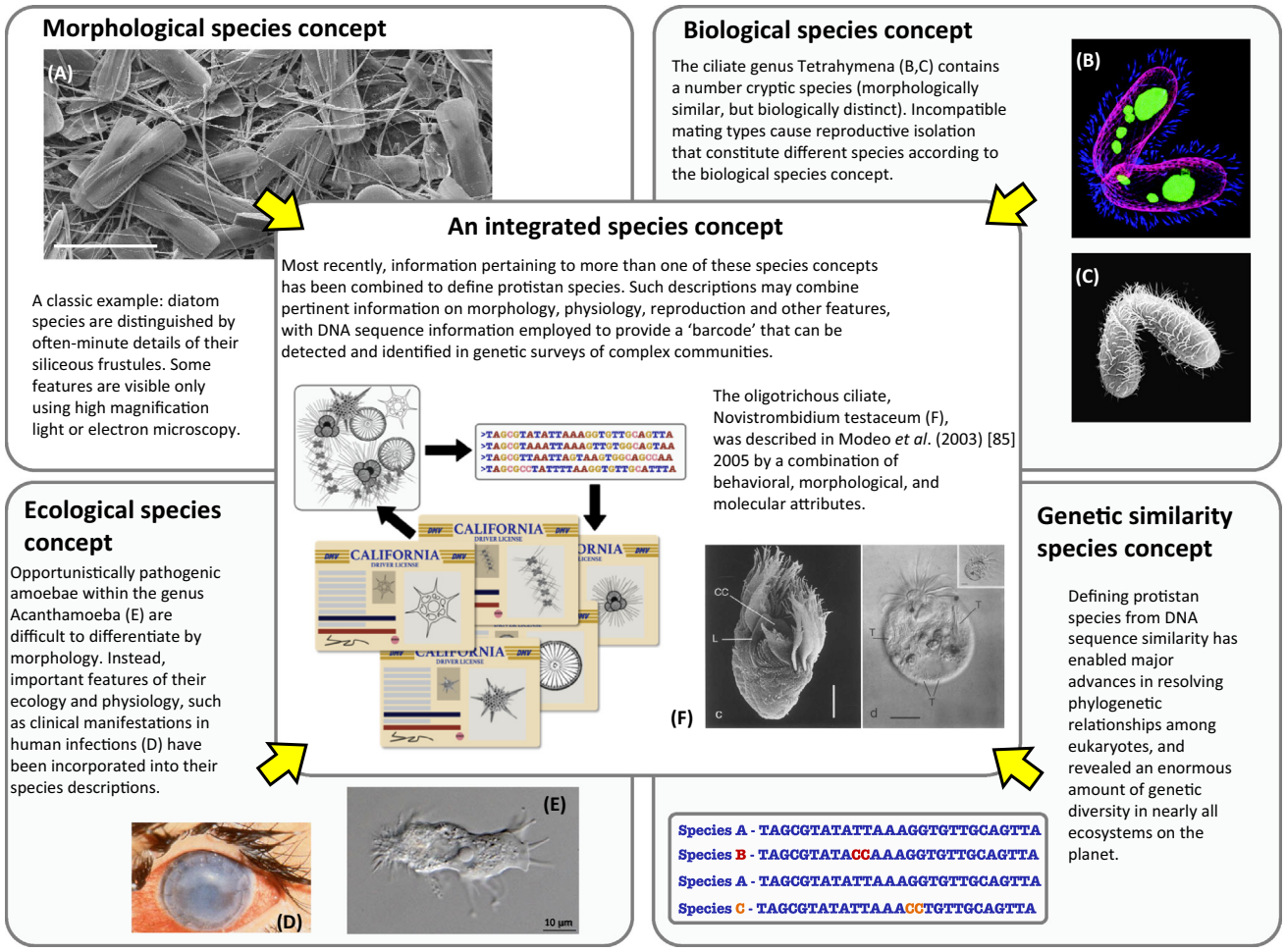
Morphology still remains the 'gold standard' for describing protistan species, but there are multiple problems applying the morphospecies concept to the study of protistan diversity. Cryptic species exist with body forms that look quite similar (i.e., indistinguishable based on morphology) but have different physiologies (i.e., they constitute different species based on an ecological species concept). Classic examples include some minute chlorophytes and chrysophytes that have few distinctive morphological features at the level of routine light microscopy but possess different nutrient-uptake capabilities, light requirements, or growth rates [12,13]. Similarly, species with amorphous morphological features, such as pathogenic amoebae, can be difficult to differentiate by morphology, thus we are forced to rely on other attributes (i.e., genetic sequence or clinical manifestations [14]) (Figure 1). Indistinguishable protistan morphotypes that contain sexually isolated mating types also confound the morphospecies concept for protists. Ciliates within the '*Tetrahymena pyriformis* complex' include several incompatible mating types, qualifying incompatible types as species according to the biological species concept [15] (Figure 1). Furthermore, some protists with complex life histories can manifest different morphologies at different life stages, resulting in the description of synonymous species [16]. These many shortcomings of the protistan morphospecies concept have made characterizing species diversity in nature quite difficult.

One practical but highly significant problem that exists for protistologists attempting to describe protistan diversity in natural communities using the morphospecies concept is the large number of protocols and taxonomic schemes that must be employed in order to identify all species. An all-encompassing taxonomic approach based on morphology is virtually impossible because of the tremendous ranges in sizes, abundances, and taxonomically informative features of protists. As a consequence, ecologists in particular have turned to DNA sequence information to develop a workable taxonomic method for assessing protistan diversity but still allowing the integration of multiple species concepts to yield species descriptions that integrate a range of information [17,18] (Figure 1).

The Double-edged Sword of Genetically Defined Protistan Species

DNA sequences have become a treasure trove of information for biologists. Sequence data have become central for addressing formerly recalcitrant questions relating to eukaryote evolution, vastly changing our view of the structure of the eukaryotic tree of life over the past 15 years. Gene-based studies of the physiological capabilities and adaptations of free-living protists using genomic and transcriptomic approaches have begun to yield fundamental information on how protists perceive and respond to changes in their chemical, physical, and biological environment [19,20].

Pertinent to this article, genetic surveys conducted on samples from natural ecosystems have yielded major advances in the way we investigate and interpret protistan species richness. Sequence-based surveys of water, sediment, and soil have resulted in the discovery of lineages of protists previously undocumented using traditional methods of microscopy and culture,



Trends in Microbiology

Figure 1. The Problematic Species Concept for Protists. Protistan species have traditionally been defined morphologically (and it is still the gold standard in the field; top left). However, lack of morphological detail for many taxa, particularly tiny ones, has resulted in the application of an integrated species concept that combines some aspects of the biological species concept (top, right), ecological species concept (lower left), and more recently species concepts based on 'genetic similarity' and phylogenetic relatedness (lower right). These disparate types of information are now being combined to yield an integrated species concept (central box). The latter effort has stimulated a tremendous effort to barcode protistan taxa, thereby enabling studies of protistan diversity, distribution, abundance, and activity. (A) A diatom assemblage collected from multiyear ice in the Ross Sea, Antarctica (marker bar = 100 μm). (B) A pair of conjugating *Tetrahymena* cells after meiosis and before nuclear exchange. Nuclei are colored green (Sytox), ciliates are blue (laboratory-made antibody), and cortex is magenta (antiserum); courtesy of Eric Cole (St Olaf College, MN, USA), adapted from Turkewitz *et al*. [82]. (C) Mating *Tetrahymena*; electron micrograph from The ASSET Program at Cornell University; courtesy of Quirk [83]. (D) *Acanthamoeba*-infected eye showing damage to the cornea and vision loss; adapted from Siddiqui and Khan [84]. (E) Light micrograph of a freshwater *Acanthamoeba*, The Netherlands; image courtesy of Ferry Siemensma. (F) Scanning electron (left) and direct interference contrast (right) micrographs of *Novistrombidium testaceum* (marker bars = 10 μm); image adapted from Modeo *et al*. [85].

including novel diversity among foraminiferans [21], marine stramenopiles [22], parasitic dinoflagellates and apicomplexans [23,24]. Genetic methods have also revealed an incredible richness of protist species in most natural ecosystems, including a tremendous array of species present at very low abundances (the 'rare protistan biosphere') that may play a role in maintaining ecosystem stability by providing ecological redundancy within microbial communities [25,26].

Genetic analyses of protistan diversity have gained widespread application and acceptance in ecology in recent years. Studies employing DNA fragment analysis or gene cloning and sequencing of 5–10 years ago have given way to high-throughput sequencing (HTS) that generates vast numbers of short-sequence reads quickly and inexpensively. DNA sequences with high similarity are generally aggregated into operational taxonomic units (OTUs). Most of these methods for assessing protistan species richness presently focus on ribosomal RNA genes, and in particular the small-subunit ribosomal RNA genes (ssu rDNA or 18S) [27,28], although barcoding projects have also examined other genes and internal transcribed spacer (ITS) regions (e.g., [29,30] and references therein).

A tacit assumption in this work is that sequence information can be organized into OTUs that roughly correspond to species-level distinctions among protists. To be sure, a molecular taxonomy of microbial eukaryotes would have significant benefits [31,32], particularly for ecologists grappling with the breadth of diversity in natural ecosystems. Unfortunately, a clear or consistent demarcation of 'species' based on sequence similarity/dissimilarity across all protistan groups has been difficult to establish, undoubtedly a consequence of limited information contained in short sequences as well as differences in the rate of evolutionary change in marker genes among the protistan lineages [33]. Values of 97–99% sequence similarity have typically been employed for forming protistan (and bacterial) OTUs from rDNA sequences, although values as low as 95% have occasionally been applied in order to minimize artificial inflation of OTU numbers due to sequencing and other errors [34–36]. Two basic approaches for grouping sequence data have been employed: *de novo* clustering, and aligning sequences to existing databases [37,38].

These rapidly advancing sequencing technologies continue to reveal more and more species richness among protists, much more than previously described or anticipated using traditional methods. For example, Finlay *et al*. [39] proposed a modest number (≈3000) for the global species richness of free-living ciliate morphotypes, while HTS approaches often detect substantially more OTUs than morphospecies [40]. This finding is not surprising because it has been established that sequence-based identities of ciliates do not necessarily agree with morphospecies assessments [41], and HTS may detect taxa present at abundances that might go undetected by microscopy [40,42]. Yet, the magnitude by which these estimates differ is alarming.

Moreover, genetic approaches actually underestimate species richness for some groups or lineages of protists, and thus the gap between traditional methods and sequence-based approaches to estimate species richness will continue to grow with advances in HTS. Some amoebozoans, rhizarians, excavates and heterolobosean amoebae are poorly represented in field surveys of 18S sequence diversity because of the poor matches of some commonly-used primers to marker genes [43,44], or especially long sequence lengths that thwart PCR amplification. These shortcomings have resulted in the use of multiplexing approaches or primer modifications [45,46] to better assess sequence diversity (and therefore protistan diversity) in samples, or group-specific primers [41] which have limited ability for estimating total protistan diversity but might better assess the diversity within a particular group [47]. As a consequence, however, differences in the specifics of primer choice, amplification protocols, and sequencing can yield rather disparate estimates of species richness based on DNA sequences, even from the same sample (an example of the variability that can be generated from a single sample set is provided in supplemental information published in [48]).

HTS has been readily adopted by ecologists despite these caveats because it has enabled scientists to conduct surveys across ambitious temporal and spatial scales. Present approaches have led to the mantra that 'revealing more diversity is better' without a clear knowledge of whether that sequence diversity is revealing species-level or intraspecific diversity (microdiversity). Haptophyte algae are a particularly interesting group in this regard, showing extreme microdiversity within their 18S genes [46,49–51]. Reports of high 'species' richness of protists are now common in the literature despite evidence that sequencing can provide estimates of richness that are sometimes orders of magnitude greater than species richness revealed by morphological studies of well known taxa such as tintinnid ciliates [52].

Bacterial ecologists have moved forward rapidly with the characterization of bacterial microdiversity, and very high estimates of gene-based bacterial species richness are common [10,37]. Their approach is sensible in that (i) many, perhaps most, free-living taxa are still unculturable, and (ii) the traditional taxonomy for bacteria is based largely on physiology. Gene sequences have been helpful in providing genetic characteristics that distinguish between bacterial 'types', that is, forms that manifest physiological differences deemed important. One of the earliest studies to apply HTS gene-based, culture-independent approaches to estimating microbial diversity was that of Sogin *et al*. [53] which examined bacterial species richness in the ocean. The authors reported bacterial species richness (OTU) estimates of >10 000 in relatively small volumes of water. This seminal paper demonstrated the potential for genetic approaches to reveal massive amounts of bacterial diversity.

It is not clear what protistologists gain from this endeavor of revealing more and more sequence diversity in nature. The bottom line is that we now accept a huge species richness as 'closer to the truth' but the relationship between sequence diversity and ecological or morphological diversity is still poorly defined. Recent studies have documented thousands of protistan 'taxa' in relatively small volumes of water, sediment, soil, or other environments [36,54–60]. Is it realistic to find this level of species richness in such small samples? It has been argued that much of the genetic diversity we see represents nothing more than neutral mutation, that is, unimportant sequence variability [61]. Certainly, this must be true at some level of sequence comparison, since we are well aware that differentiating virtually every human on the planet is possible with sufficient DNA sequencing (yet we are clearly one species if applying the biological species concept). By analogy, it would appear that protistan ecologists have arrived at the point where we have moved past our inability to catalog species richness using traditional methods to face the difficult task of separating uninformative sequence differences (neutral mutations) from informative ones, the latter being those differences that correspond to features of protists that we deem sufficiently important to confer unique scientific names (e.g., differences in morphology, physiology, or mating types).

Objectives for Assessing Protistan Species Diversity in the Age of HTS

Protistologists have grappled for centuries attempting to document the diversity of single-celled eukaryotes in natural ecosystems using morphological and physiological criteria, with limited success. Given present and emerging sequencing capabilities, the task before protistan ecologists at this time and into the future will be focused on making sense of the enormity of DNA sequence datasets that are becoming available for protists, and melding that information with traditional identifications (Figure 1). In some situations this may amount to accepting fairly large sequence variability (e.g., 95%) within traditionally defined protistan taxa, while in other cases much higher resolution may actually be desirable. In all cases the approach will be taxon- and goal-dependent.

Define and Measure the 'Appropriate' Level of Sequence Diversity

We clearly now possess the potential (some might argue the actuality) of assessing protistan species richness in nature far beyond species-level distinctions using modern genetic approaches, and in the process generating a tremendous amount of information that has no practical (i.e., ecological) value. It has been estimated that there are $>10^{30}$ microbes on the planet [62]. Our goal as ecologists should not simply be the endless gathering and splitting of sequence data into myriad OTUs without ecological context or usefulness. It would appear that some researchers have lost sight of the need for ecological context in their fervor to characterize more and more genetic diversity. It is noteworthy that the diversity of form and function among protists that we can presently define (without genetic approaches) is already overwhelming for modelers of global biogeochemistry. More diversity is not necessarily better. Should we continue to characterize more and more genetic diversity because we can?

There is no simple answer to that question because there can be valid reasons for characterizing genetic microdiversity (i.e., intraspecific sequence variation within morphospecies) if it provides additional criteria for identifying important capabilities. For example, Martens *et al*. [63] noted different analogs and amounts of saxitoxins and spiroimine toxins produced among 68 isolated strains of the dinoflagellate *Alexandrium ostenfeldii*. Differences in toxin production in this dinoflagellate transcend the morphospecies concept. Whether these *A. ostenfeldii* strains should now be described as many different species or accepted as variants within a single species is a subject for taxonomists to debate. Defining each strain as a different species would fly in the face of long-standing consensus that all populations and species exhibit some level of physiological variability [64,65], although, from a human health perspective, investigating their genetic microdiversity might reveal key genetic markers indicative of toxin production (and therefore be highly useful for ecologists and toxicologists).

Such vexing complications accepted, one can reasonably ask in a general sense if we have gone rapidly from 'not seeing everything' to 'seeing too much' with genetic assessments of species richness among protists. Paradoxically, some recent estimates of the number of protistan OTUs are now approaching the total number of individual protistan cells that have been observed and counted by microscopy in samples of similar size, but OTU rarefaction curves constructed from these same datasets indicate that more OTUs remain to be detected. These findings imply that we are rapidly moving past the point of being unable to 'see' the species richness of natural protistan communities to the point of characterizing strain-level differences within species (i.e., microdiversity). The key for ecologists utilizing sequence information is establishing the level of richness that is appropriate for their scientific question or goal.

Improve Criteria for Forming Protistan OTUs

The criteria presently used by protistologists to form protistan OTUs have been a constantly moving target. The inconsistency of generating OTUs [48] has thwarted comparisons of protistan richness estimates across studies since these approaches first appeared. There may be internal consistency for some datasets (e.g., within a laboratory or working group), but present methods for forming OTUs vary in many ways, including: sampling/filtration, template storage, extraction protocol, gene target/region, primer choice and amplification protocol, library preparation, sequencing platform, error filtering and chimera checking, and OTU-calling parameters. Consistency is highly desirable in order to facilitate global (or local) assessments of protistan species richness, and species distributions. Directly related to this goal, sequence information will continue to expand rapidly, revealing more genetic diversity within protistan communities. A path forward is needed that differentiates interspecies from intraspecies sequence variability for various protistan groups, including strict quality control on sequences and some level of standardization for OTU calling.

Fortunately, centralized databases and standardized approaches for forming and naming OTUs [66–69] and comparing disparate datasets [70,71] are beginning to address the issue of noncomparability among studies of protistan diversity [17]. Improvements in denoising and chimera detection [72,73], clustering approaches to form OTUs [38,74], and evaluation of evolutionary models and diversity indices used to estimate richness [75] continue to appear in the literature at a rapid rate. The use of mock communities, well defined assemblages of species that can be used to vet HTS applications that can then be applied to natural assemblages of microbial eukaryotes, will also improve our use of these powerful albeit often poorly understood methods [76]. Moreover, many investigators now acknowledge the shortcomings of OTU calling (specifically the lack of consistency/comparability), and alternative methods have been proposed. Amplicon sequence variants (ASVs) [77] and oligotyping [78,79] provide alternative approaches for examining sequence-based diversity that may generate outcomes that are more comparable across studies [48], yet appear to yield ecological patterns that are similar to those generated by OTUs [80]. We should encourage such new perspectives on sequence diversity, and strive to understand their meaning within the context of protistan morphology and physiology.

Outstanding Questions

Have we moved rapidly past the point of being unable to describe the species-level richness of natural protistan communities to the point of characterizing strain-level differences within traditional morphospecies (i.e., microdiversity)?

If so, what are the benefits and consequences to ecologists who might want to continue to plumb the depths of protistan diversity beyond the species level?

Concluding Remarks

Should ecologists continue to plumb the depths of protistan diversity beyond the species level using DNA sequences? The answer to that question is 'yes', but the work should be carried out with full recognition and disclosure. There is now mounting information that strain-to-strain variability is an important component of the ecology of many protistan morphospecies, warranting genetic characterizations that attempt to resolve important physiological differences contained within them. Ecologists must embrace a more nuanced (and more complicated) view of the species concept for protists (see Outstanding Questions) [81]. Conversely, DNA sequence diversity means nothing if that diversity does not differentiate protistan 'types' by criteria that correlate to differences in morphology, function, secondary metabolites, behavior, mating type, or some other feature(s) deemed important. Simply defining more and more species richness should not, by itself, be a goal of our molecular methods in protistan ecology. Defining meaningful diversity should be.

References

1. Armbrust, E. (2009) The life of diatoms in the world's oceans. *Nature* 459, 185–192
2. Worden, A.Z. *et al.* (2015) Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* 347, 6223
3. Caron, D.A. and Goldman, J.C. (1990) Protozoan nutrient regeneration. In *Ecology of Marine Protozoa* (Capriulo, G.M., ed.), pp. 283–306, Oxford University Press
4. Geisen, S. *et al.* (2017) Soil protistology rebooted: 30 fundamental questions to start with. *Soil Biol. Biochem.* 111, 94–103
5. Xiong, W. *et al.* (2017) Soil protist communities form a dynamic hub in the soil microbiome. *ISME J.* 12, 634
6. Armstrong, H.A. and Brasier, M.D. (2013) *Microfossils,* Blackwell
7. Burki, F. (2014) The eukaryotic Tree of Life from a global phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* 6, a016147
8. Simpson, A.G.B. *et al.* (2017) Protist diversity and eukaryote phylogeny. In *Handbook of the Protists* (Archibald, J.M., ed.), pp. 1–21, Springer
9. Corliss, J.O. (1975) Three centuries of protozoology: a brief tribute to its founding father, A. van Leeuwenhoek of Delft. *J. Protozool.* 22, 3–7
10. Locey, K.J. and Lennon, J.T. (2016) Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. U. S. A.* 113, 5970–5975
11. Zinger, L. *et al.* (2012) Two decades of describing the unseen majority of aquatic microbial diversity. *Mol. Ecol.* 21, 1878–1896
12. Bock, C. *et al.* (2017) Genetic diversity in chrysophytes: comparison of different gene markers. *Fottea* 17, 209–221
13. Fawley, M.W. *et al.* (2006) Evaluating the morphospecies concept in the Selenastraceae (Chlorophyceae, Chlorophyta). *J. Phycol.* 42, 142–154
14. Gast, R.J. *et al.* (1996) Subgenus systematics of *Acanthamoeba*: four nuclear 18S rDNA sequence types. *J. Eukaryot. Microbiol.* 43, 498–504
15. Simon, E.M. *et al.* (2008) The '*Tetrahymena pyriformis*' complex of cryptic species. *Biodivers. Conserv.* 17, 365–380
16. Cedhagen, T. and Tendal, O.S. (1989) Evidence of test detachment in *Astrorhiza limicola*, and two consequential synonyms: *Amoeba gigantea* and *Megamoebomyxa argillobia* (Foraminiferida). *Sarsia* 74, 195–200
17. Berney, C. *et al.* (2017) UniEuk: time to speak a common language in protistology! *J. Eukaryot. Microbiol.* 64, 407–411
18. del Campo, J. *et al.* (2018) EukRef: phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. *PLoS Biol.* 16, e2005849

19. Caron, D.A. *et al.* (2016) Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat. Rev. Microbiol.* 15, 6–20
20. Sibbald, S.J. and Archibald, J.M. (2017) More protist genomes needed. *Nat. Ecol. Evol.* 1, 0145
21. Pawlowski, J. *et al.* (2014) Next-generation environmental diversity surveys of foraminifera: preparing the future. *Biol. Bull.* 227, 93–106
22. Massana, R. *et al.* (2004) Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl. Environ. Microbiol.* 70, 3528–3534
23. Horiguchi, T. *et al.* (2015) Diversity and phylogeny of marine parasitic dinoflagellates. In *Marine Protists: Diversity and Dynamics* (Ohtsuka, S., ed.), pp. 397–419, Springer
24. Mahé, F. *et al.* (2017) Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat. Ecol. Evol.* 1, 0091
25. Caron, D.A. and Countway, P.D. (2009) Hypotheses on the role of the protistan rare biosphere in a changing world. *Aquat. Microb. Ecol.* 57, 227–238
26. Logares, R. *et al.* (2014) Patterns of rare and abundant marine microbial eukaryotes. *Curr. Biol.* 24, 813–821
27. Dunthorn, M. *et al.* (2012) Comparing the hyper-variable V4 and V9 regions of the small subunit rDNA for assessment of ciliate environmental diversity. *J. Eukaryot. Microbiol.* 59, 185–187
28. Hu, S.K. *et al.* (2015) Estimating protistan diversity using high-throughput sequencing. *J. Eukaryot. Microbiol.* 62, 688–693
29. Hebert, P.D.N. *et al.* (2003) Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* 270, 313–321
30. Pawlowski, J. *et al.* (2012) CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.* 10, e1001419
31. Caron, D.A. (2013) Towards a molecular taxonomy for protists: benefits, risks and applications in plankton ecology. *J. Eukaryot. Microbiol.* 60, 407–413
32. Bucklin, A. *et al.* (2016) Metabarcoding of marine zooplankton: prospects, progress and pitfalls. *J. Plankton Res.* 38, 393–400
33. Nebel, M. *et al.* (2011) Delimiting operational taxonomic units for assessing ciliate environmental diversity using small-subunit rRNA gene sequences. *Environ. Microbiol. Rep.* 3, 154–158
34. Caron, D.A. *et al.* (2009) Defining DNA-based operational taxonomic units for microbial eukaryote ecology. *Appl. Environ. Microbiol.* 75, 5797–5808
35. Debroas, D. *et al.* (2015) Evidence for an active rare biosphere within freshwater protists community. *Mol. Ecol.* 24, 1236–1247
36. Xu, D. *et al.* (2017) Microbial eukaryote diversity and activity in the water column of the South China Sea based on DNA and RNA high throughput sequencing. *Front. Microbiol.* 8, 1121
37. Edgar, R.C. (2017) Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. *PeerJ* 5, e3889
38. Forster, D. *et al.* (2016) Comparison of three clustering approaches for detecting novel environmental microbial diversity. *PeerJ* 4, e1692
39. Finlay, B.J. *et al.* (1998) Protozoan diversity: converging estimates of the global number of free-living ciliate species. *Protist* 149, 29–37
40. Santoferrara, L.F. *et al.* (2016) Patterns and processes in microbial biogeography: do molecules and morphologies give the same answers. *ISME J.* 10, 1779–1790
41. Stoeck, T. *et al.* (2014) A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. *Environ. Microbiol.* 16, 430–444
42. Santoferrara, L.F. *et al.* (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
43. Hugerth, L.W. *et al.* (2014) Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. *PLoS One* 9, e95567
44. Jeon, S.-O. *et al.* (2008) Environmental rRNA inventories miss over half of protistan diversity. *BMC Microbiol.* 8, 222–234
45. Stoeck, T. *et al.* (2006) A multiple PCR-primer approach to access the microeukaryotic diversity in environmental samples. *Protist* 157, 31–43
46. Balzano, S. *et al.* (2015) Protist diversity along a salinity gradient in a coastal lagoon. *Aquat. Microb. Ecol.* 74, 263–277
47. Doherty *et al.* (2007) Culture independent assessment of planktonic ciliate diversity in coastal northwest Atlantic waters. *Aquat. Microb. Ecol.* 48, 141–154
48. Hu, S.K. and Caron, D.A. (2018) Are we overestimating protistan diversity in nature? *Zenodo* Published online October 2, 2018. http://dx.doi.org/10.5281/zenodo.1442685
49. Liu, H. *et al.* (2009) Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12803–12808
50. Needham, D.M. *et al.* (2017) Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J.* 11, 1614–1629
51. Egge, E. *et al.* (2013) 454 pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. *PLoS One* 8, e74371
52. Bachy, C. *et al.* (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
53. Sogin, M.L. *et al.* (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12115–12120
54. Chen, W. *et al.* (2017) Patterns and processes in marine microeukaryotic community biogeography from Xiamen coastal waters and intertidal sediments, Southeast China. *Front. Microbiol.* 8, 1912
55. Tragin, M. *et al.* (2018) Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. *Environ. Microbiol.* 20, 506–520
56. de Vargas, C. *et al.* (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* 348, 1261605
57. Pernice, M.C. *et al.* (2015) Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J.* 10, 945
58. Ishaq, S.L. *et al.* (2017) An investigation into rumen fungal and protozoal diversity in three rumen fractions, during high-fiber or grain-induced sub-acute ruminal acidosis conditions, with or without active dry yeast supplementation. *Front. Microbiol.* 8, 1943
59. Debroas, D. *et al.* (2017) Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available metabarcoding data. *FEMS Microbiol. Ecol.* 93, fix023
60. Nolte, V. *et al.* (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Mol. Ecol.* 19, 2908–2915
61. Fenchel, T. (2005) Cosmopolitan microbes and their 'cryptic' species. *Aquat. Microb. Ecol.* 41, 49–54
62. Kallmeyer, J. *et al.* (2012) Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc. Natl. Acad. Sci. U. S. A.* 109, 16213–16216
63. Martens, H. *et al.* (2017) Toxin variability estimations of 68 *Alexandrium ostenfeldii* (Dinophyceae) strains from the Netherlands reveal a novel abundant Gymnodimine. *Microorganisms* 5, 29
64. Brand, L.E. (1981) Genetic variability in reproduction rates in marine phytoplankton populations. *Evolution* 35, 1117–1127
65. Burkholder, J.M. and Glibert, P.M. (2009) The importance of intraspecific variability in harmful algae – Preface to a collection of topical papers. *Harmful Algae* 8, 744–745
66. Guillou, L. *et al.* (2012) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* 41 DI, D597–D604
67. Caporaso, J.G. *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336

68. Morard, R. *et al.* (2015) PFR2: a curated database of planktonic foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography and evolution. *Mol. Ecol. Res.* 15, 1472–1485
69. Mordret, S. *et al.* (2018) DINOREF: A curated dinoflagellate (Dinophyceae) reference database for the 18S rRNA gene. *Mol. Ecol. Res.* 18, 974–987
70. Massana, R. *et al.* (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environ. Microbiol.* 17, 4035–4049
71. Grossmann, L. *et al.* (2016) Protistan community analysis: key findings of a large-scale molecular sampling. *ISME J.* 10, 2269–2279
72. Bokulich, N.A. *et al.* (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* 10, 57–59
73. Callahan, B.J. *et al.* (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581
74. Kopylova, E. *et al.* (2016) Open-source sequence clustering methods improve the state of the art. *mSystems* 1, e00003-15
75. Barley, A.J. and Thomson, R.C. (2016) Assessing the performance of DNA barcoding using posterior predictive simulations. *Mol. Ecol.* 25, 1944–1957
76. Geisen, S. *et al.* (2015) Not all are free-living: high-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. *Mol. Ecol.* 24, 4556–4569
77. Callahan, B.J. *et al.* (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 11, 2639
78. Eren, A.M. *et al.* (2014) Oligotyping analysis of the human oral microbiome. *Proc. Natl. Acad. Sci. U. S. A.* 111, E2875–E2884
79. Eren, A.M. *et al.* (2014) Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high- throughput marker gene sequences. *ISME J.* 9, 968–979
80. Glassman, S.I. and Martiny, J.B.H. (2018) Ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. *bioRxiv* Published online March 15, 2018. http://dx.doi.org/10.1101/283283
81. Mulot, M. *et al.* (2017) Genetic determinism vs. phenotypic plasticity in protist morphology. *J. Eukaryot. Microbiol.* 64, 729–739
82. Turkewitz, A.P. *et al.* (2002) Functional genomics: the coming of age for *Tetrahymena thermophila*. *Trends Genet.* 18, 35–40
83. Quirk, T. (2013) How a microbe chooses among seven sexes. *Nature* Published online March 27, 2013. http://dx.doi.org/10.1038/nature.2013.12684
84. Siddiqui, R. and Khan, N.A. (2012) Biology and pathogenesis of *Acanthamoeba*. *Parasites Vectors* 5, 6
85. Modeo, L. *et al.* (2003) A multidisciplinary approach to describe protists: redescriptions of *Novistrombidium testaceum* and *Strombidium inclinatum* Montagnes, Taylor and Lynn 1990 (Ciliophora, Oligotrichia). *J. Eukaryot. Microbiol.* 50, 175–189