Review

Understanding marine microbial eukaryotes in a genomic age means accepting that protists are not just big bacteria

(Au: This title is quite long. Could you try to come up with something more succinct? Maybe just "Understanding marine microbial eukaryotes in a genomic age"?)

Patrick J. Keeling and Javier del Campo

Botany Department, University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC, V6T 1Z4, Canada.

E-mail: pkeeling@mail.ubc.ca

The study of marine microbial ecology has been completely transformed by molecular and genomic data: after centuries of relative neglect, genomics has revealed the surprising extent of microbial diversity and how microbial processes transform ocean and global ecosystems. But the revolution is not complete: major gaps in our understanding remain, and one obvious example is that microbial eukaryotes, or protists, are still largely neglected. Here we examine various ways in which protists might be better integrated into models of marine microbial ecology, what challenges this will present, and why understanding the limitations of our tools is a significant concern. In part this is a technical

challenge — eukaryotic genomes are more difficult to characterize — but eukaryotic adaptations are also more dependent on morphology and behaviour than they are on the metabolic diversity that typifies bacteria, and these cannot be inferred from genomic data as readily as metabolism can be. We therefore cannot simply follow in the methodological footsteps of bacterial ecology and hope for similar success. Understanding microbial eukaryotes will require different approaches, including greater emphasis on taxonomically and trophically diverse model systems. Molecular sequencing will continue to play a role, and advances in environmental sequence tag studies and single-cell methods for genomic and transcriptomics offer particular promise.

Introduction

In *Essay Concerning Human Understanding* (1689), John Locke wrote, “It is of great use to the sailor to know the length of his line, though he cannot with it fathom all the depths of the ocean. It is well he knows that it is long enough to reach the bottom at such places as are necessary to direct his voyage, and caution him against running upon shoals that may ruin him.” Locke was writing about understanding the tools of human thought, but it is every bit as sensible to understand the limits of those tools that contributed to the creation of a body of data as well, because these affect our interpretation every bit as acutely. The metaphor is especially apt when applied to the tangled web of networks that makes up marine microbial ecosystems: here is a problem where the uncharted waters still greatly outweigh our understanding, and yet we are suddenly moving so astonishingly quickly that the raw information available greatly

exceeds our conscious appreciation of both the strengths and weaknesses of many tools used to create and analyse this information, or how they affect our interpretation of it.

The problem is complex, but important, because microbial life drives every one of the major biogeochemical cycles that make the ocean so central to all other ecosystems on earth [1]. We and our animal and plant cousins depend on these cycles and sometimes also have strong effects on them — pulling them in one direction or another by our activities — but the engines driving nutrient and energy flow are all microbial processes [2]. Because microbial communities form the foundations of these ecosystems, disrupting them can also have profound impacts on the rest of the system, much in the same way that shifting the foundation of a tall building may be amplified through the structure in ways we cannot see until it is too late to mitigate against the damage.

Despite their impact on a planetary scale, we know relatively little about the composition of many of these microbial communities and even less about how they interact and function at the ecosystem level. Indeed, we probably know more about some individual fish or marine mammalian species as we do, collectively, about the tens of thousands of microbial species whose activities allow those larger and more charismatic species to survive. The reasons for this are manifold, but broadly trace back to two simple problems: microbes are small, and microbes are diverse. The first of these probably should not come as a surprise, but the microbial world is *so* small that in many ways it lays even beyond our imagination, and size poses challenges both of technical nature and in failing to command our attention. For a microbial ecologist studying

marine bacteria, for example, the scale of the organism makes observation somewhat like a traditional ecologist trying to study animals on the African savanna from space: the vastness of the size difference makes the problem qualitatively different. Because we can't see them, we also tend to ignore microbes, and observing their effects demands greater attention, not less.

The second challenge is the vastness of the biological diversity that we lump together with tags like 'microplankton', 'nanoplankton', 'picoplankton', or even 'the microbial world' [1]. Historically, microbes have been relegated to an intellectual stew: it was difficult to see much diversity with available tools, and what we could see was even more difficult to interpret [3]. Bacterial diversity is largely not manifested at the level of structure, but rather at the level of molecules, particularly metabolism. Bacterial morphology afforded few characters that allowed us to infer their evolutionary relationships, but their metabolic diversity is so extreme that it had the opposite problem: metabolic variation evolves so rapidly that patterns of that variation shed little light on how bacteria were related to one another or to eukaryotes, or even how much evolutionary diversity bacteria encompass. Microbial eukaryotes, by contrast, exhibit a great deal of morphological diversity (e.g., see Figure 1), but even once the tools to look at the relevant scale were available (e.g. light and electron microscopy) it also proved too much to reveal patterns to help interpret this diversity or meaningfully relate it to macroscopic life [4]. Moreover, only a tiny proportion of protists and prokaryotes are readily cultivable, so together our inability to identify them, interpret their diversity, or cultivate them to study them in detail all contributed to the lumping of microbes into functional classes so broad as to be effectively meaningless [5–7].

The breakthrough for understanding the scope of microbial diversity was molecular biology, and phylogenetic trees based on molecular sequence data in particular. Molecular trees gave us our first real view of just how diverse the microbial world is, and how microbes are related to more familiar animals, fungi, and plants. The view was dramatic, but also problematic: in molecular trees, microbial diversity outstrips all other biodiversity by orders of magnitude: tiny creatures we thought were the same species turned out to be as distant from one another as humans are from chickens, or in some cases as humans are from pine trees [8]. This change is largely a matter of perspective (things we think are ‘important’ distinctions are just our opinions), but is still an important problem when it comes to relating the somewhat more abstract diversity of microbial life to specialists outside the field, to whom microbes can still appear to lack diversity.

Molecular phylogeny opened up the ‘black box’ of microbial diversity and offered ways to manage it. The field of microbial ecology is now rapidly growing and our understanding is moving faster than ever, largely based on using molecular tools to study microbial diversity in the field, and molecular phylogenetic frameworks to interpret that data [1,2,6,7]. Communities of uncultivated microbes can be examined collectively using environmental ‘tag’ sequencing, where a fragment of the small subunit ribosomal RNA (SSU rRNA) is sampled from an environment to quickly identify most inhabitants. Similarly, various kinds of ‘meta-omics’ can tell us what genes are possessed by communities as a whole (metagenomics), which genes are being expressed (e.g., metatranscriptomics) [9,10] or proteins being translated (metaproteomics) [11]. Advances have been made, and more are on the horizon, but before we celebrate the

endless bounty of a new age of molecular microbial ecology, we should pause and consider the microbial eukaryotes.

Virtually every microbial ecosystem known in the oceans, in freshwater, and on land, includes eukaryotes (the exceptions being some of the most 'extreme' environments), where they play some of the same roles as bacteria and archaea, but also have some essential and unique ecological roles [12]. But, just as the microbial world as a whole was overshadowed for centuries by our concentration on animals and plants, our current understanding of microbial ecosystems concentrates strongly on bacteria and archaea, overshadowing the roles of eukaryotes (Figure 2). Below, we discuss some of the challenges we face in more adequately integrating microbial eukaryotes into our understanding of microbial ecosystems in general, and whether we adequately know 'the length of our line' to steer this voyage from the shoals.

M**orphology and Behaviour Fundamentally Change How We Must Make Functional Predictions From Eukaryotic Genomes**

There are numerous reasons why we know comparatively less about microbial eukaryotes than we do about bacteria, even when they serve important functions in otherwise well-studied ecosystems. Some of these reasons reflect mundane quantitative differences: eukaryotes tend to have smaller populations and larger genomes with larger families of genes [13]. These pose challenges to bulk sequencing approaches, although theoretically solvable through more sequencing. Of course, nuclear genomes are not just larger, but also more complex: multiple chromosomes,

repeats, vast stretches of non-coding sequence, spliceosomal introns, and variations in sequence composition within a genome all hamper the interpretation of nuclear genomic data. At the extreme end of the spectrum, some genomes are so large as to defy any known correlation with other aspects of the cell's biology (e.g. some amoebae, dinoflagellates, or euglenozoans [14]), and fully assembling these genomes is likely still beyond current sequencing technology under ideal conditions, much less when sampled in a soup of nuclear metagenomics.

However, there are also qualitative differences between prokaryotes and eukaryotes that lead to problems that are at once more challenging, but also more biologically interesting. First and foremost, we argue that eukaryotes and prokaryotes principally depend on different evolutionary strategies to adapt to ecological niches, which leads to the manifestation of biodiversity at different levels [1,15]. To generalize, bacteria are morphologically relatively simple and homogeneous, but at the level of metabolism they display a wondrous level of diversity. Bacteria have been found that can thrive in almost any set of physical conditions, and to extract energy and nutrients from almost any chemical and bioenergetic scenario [16]. In contrast, eukaryotes are generally more metabolically simple and homogeneous, but at the level of morphology are wondrously diverse [17]. In addition to behaviours at the molecular level that are also common in prokaryotes (like sensing constituents of the environment), eukaryotic microbes have a multitude of structural innovations that give them a greater capacity for diverse means to achieve active behavioural complexity: activities such as defence and attack, infection, modes of locomotion, feeding, reproduction, buoyancy control, and so on are effected by many strategies that rely on structural and behavioural diversity.

These are generalizations of course; bacteria do not lack morphological diversity or complex behavioural characteristics (just watch *Bdellovibrio* attacking another bacterium [18]), just the same as eukaryotes do not lack metabolic diversity (consider the complexity of fungal metabolism [19]). But in general, prokaryotes and eukaryotes rely to different degrees on these different strengths for adaptation.

This generalization also has important technical implications for how we approach the problem of fitting eukaryotes into microbial ecology data and theory. The great leap forward for microbial ecology based on 'omics' owes its remarkable success to our ability to accurately identify metabolic functions from gene sequence data, which is itself due to the relatively simple modularity of metabolic proteins. Assuming our functional identification of metabolic genes is largely correct (an important assumption to examine, but one that we will set aside here), we have fundamental insights into how these enzymes work that allows us to transfer those insights to new species as they are investigated based on the further assumption that a given enzyme will perform a similar reaction when plugged into an otherwise coherent metabolic network. This gives us great power to predict functional roles and interaction networks for bacteria, because those networks are primarily based on metabolism [20] (bacterial networks have many other levels, but study of their ecology has largely focused on metabolic networks). Indeed, environmental microbiologists working on bacteria are already verifying these predicted functions and interactions *in situ* through the use of metaproteomics and metabolomics [21,22]. The same well-tested genomic tools, environmental sequence tags, metagenomics, and metatranscriptomics, are now being applied to microbial

eukaryotes, but is there any expectation that they will be as successful? We argue there are good reasons to think not [15].

To illustrate this question with something less abstract than a protist, consider the case of the New Caledonian Crow. This bird is known for its intelligence: it uses multi-step problem solving and understands and uses abstract phenomena such as displacement. Ask yourself, given its complete genome sequence but no other information, could we infer its intelligence and problem solving characteristics? The answer of course is ‘no’. Indeed, we argue further that without the benefit of a great deal of context we would not even be able to conclude that it had feathers and could fly, because there is little in the genome to immediately suggest ‘birdness’ either. This is because we cannot readily translate gene sequences into complex multi-gene traits where much of the variation comes from expression levels and regulation networks rather than more readily interchangeable modules like enzymes, and virtually everything to do with structures and behaviour tends to be based upon such traits. Or in other words, we lack the equivalent of the fundamental knowledge of enzyme function that allows us to interpret metabolic genes in bacteria, and without that we cannot make such detailed ecological and functional inferences from eukaryotic genomes as we can for bacteria. Indeed, for eukaryotic traits of interest it is even possible that any equivalent fundamental understanding will never exist. This is because the characteristics appear to evolve more contingently and more often convergently. In this case, similar ecological strategies can evolve without a fundamentally homologous basis, and these would be impossible to infer from one genome based on information gleaned from another genome where a non-homologous system evolved for the same

purpose. For example, assume that the structural basis for saprotrophy in oomycetes and fungi may have evolved in each group independently from non-homologous components (or from non-specific cellular components that have roles in many systems). If so, then this feeding mechanism could not be inferred from the genome of one lineage even with a perfect understanding of the molecular basis of the system in the other lineage. In cases where complex morphology and behaviour are underpinned by fundamentally homologous traits, there is hope that a detailed understanding of model systems will one day allow us to infer these characteristics in other organisms based on the genome alone, but even then the inference will be much more complex than it is for metabolic pathways in bacteria.

While microbial eukaryotes may be small and have overlapping ecological roles with bacteria, they are still eukaryotes after all, and the same basic problem extends just as much to a dinoflagellate as to a bird. Even with complete nuclear genome sequences, therefore, our conclusions about ecological roles of protists can be limited to ridiculously broad conclusions, like ‘it’s photosynthetic’ [15]. We lack genomic flags for even the broadest of defined ecological roles such as ‘parasitic’, or ‘predator’, and most likely such flags simply do not exist [23], so it is not just a matter of learning to recognize them. Worse still, the complexity of eukaryotic behaviour means that there is overlap within a single species between the kinds of roles we like to identify: for example, many or most ‘photosynthetic’ protists are actually better described as mixotrophs, because they are also heterotrophic feeders, sometimes eating more bacteria than the purely heterotrophic protists in the same environment [24].

Thankfully, a solution to this problem does not require a great leap of logic or theory, because the solution is already common practice, albeit imperfectly applied. We routinely make assumptions about what a little-studied species is like and what its role in the environment is likely to be based on what its close relatives are like and do. For example, with a crow genome we would quickly infer it had feathers and could fly because it was obvious from the sequence that it was related to other feathered, flying animals, rather than because we identified genes related to feathers or flying. This comparative reasoning can be very powerful, but also depends heavily on the quality and quantity of information one has from those close relatives.

The first such requirement is a reference tree, which is necessary to correctly identify the subject's closest relatives. We currently lack a well-supported reference tree spanning all microbial eukaryotes, but we have most of the data one would need to develop it, and this is a relatively straightforward problem to solve that is currently underway for the most commonly used tag sequence, SSU rRNA [25]. The second requirement is for a large body of information about the biology of closely related species. Obviously, you learn little about a species by identifying a closely related species about which we also know nothing. Moreover, the closer the relationship, the more likely the inferences are to be true; our inferences about the crow might be misleading if the nearest relative we could compare it to was a crocodile.

This is a considerably more difficult problem that does not readily lend itself to 'high-throughput' solutions, because what this really means is we need a large number of relatively well-studied model organisms scattered around each and every major branch of the tree of eukaryotes. We need to go back to nature and actually look at how

these cells work and what they are doing in the environment. By developing more model systems distributed through the tree of eukaryotes, and across different ecological roles, we will massively improve our ability to infer ecological roles of environmental samples because our assumptions about roles based on sequence data will be much more accurate [26,27].

What Can We Learn From Environmental Tag Sequencing?

As with bacteria, the first clone library studies of microbial eukaryotes revealed a picture of the marine diversity that was different form the picture sketched out over the previous decades by microscopy [28,29]. Two group of picoeukaryotes in particular stood out: the parasitic MALV (Marine Alveolates, or syndinians) and the bacterivorous MAST (Marine Stramenopiles) lineages emerged as some of the most abundant organisms in the sea, representing up to 50% of the sequences in seminal molecular environmental protistolgy studies [30]. In addition to these two groups, a myriad of other branches sprouted from as many parts of the eukaryotic tree, corresponding to additional novel, uncultured lineages [31]. High-throughput tag sequencing methods have now replaced clone libraries for the study of protist diversity in the environment [32,33] and further accelerated these discoveries.

But to make full use of these data, our ability to interpret them needs to keep pace with data generation. The reference data discussed above forms the foundation for how we interpret molecular tag sequencing from environmental samples, and we need to understand how well the tags reflect the composition of the communities that

they are meant to represent. As with bacteria, the current gene of choice to survey eukaryotic diversity is the SSU rRNA. Next generation sequencing methods require short fragments, so two different regions of the SSU have been used to study eukaryotic diversity, V4 and V9 [34]. Each has known biases, some of which were predictable based on the sequences of these regions: for example, V4 excludes euglenozoans [35], and so the diversity of marine diplonemids (a subgroup of euglenozoans) was not revealed until large surveys using V9 were carried out [36]. Another major bias relates to the amplification of sequence tags versus directly isolating cells from the environment. When libraries of amplified and cloned SSU fragments were compared with a database of the same SSU fragment taken from large numbers of sorted single cells from the same environment, the proportions of some taxa were similar, but others were significantly different [37]. Some of these differences related to particular taxa being over-represented in the amplified tags, but other differences were more general and crossed taxonomic lines, the most significant being the under-representation of heterotrophs as a whole in clone libraries. These biases are attributed to the copy number of SSU rRNA genes in different genomes [37,38]. While bacteria have between one and ten copies of the SSU in their genomes, microbial eukaryotes can have a number of copies that differ by many orders of magnitudes. Some groups, like dinoflagellates, can have up to 12,000 copies [38] while others, like MAST-4, have only 30 [39]. Difference in the SSU copy number contribute to ‘abundance’ when analyzing DNA- or RNA-derived tag data or metagenomes [35,40].

Other less obvious biases probably also exist, and understanding these and how they might affect major patterns is a key problem. One approach to identifying cryptic

biases is to look for unrecognized patterns in the data and then ask whether they reveal biological phenomena or methodological biases. Normally such data are broken down by environmental criteria, but one can look at patterns *a priori*, across taxa and between different methods and markers. As a simple example, here we have re-analysed the complete set of microbial eukaryotic tag V9 sequences from Tara Oceans [41], the largest database of eukaryotic tags currently available. We examined the relative abundance of the 25 most common individual operational taxonomic units (OTUs; e.g., 'species') from the whole dataset, as well as the relative abundance of the 10 most common OTUs within 12 well-defined and diverse protist lineages (dinoflagellates, radiolarians, diatoms, diplonemids, fungi, pelagophytes, syndinians, apicomplexans, ciliates, green algae, and bygirans, which are heterotrophic stramenopiles). Across the whole data set, we find that only 8 OTUs represent 50% of all reads (Figure 3). Taken at face value, this means a small number of species are hyper-abundant, which fits with the expected distribution [42]. Interestingly, however, when we look at the structure of the read distribution within each of the 12 common and diverse lineages, we observe that not all the groups share the same pattern. Instead, we see two different types of distributions — normal vs jackpots. In the first instance, the ten most abundant OTUs account for much of the data, but we see a gradual decline in relative abundance and a significant proportion of the reads from the entire group are distributed across less abundant OTUs. This is seen for the radiolarians, haptophytes, ciliates, green algae, bygirans (among them some of the abundant MAST groups), and the fungi, where the most common taxa are abundant, but no one taxon dominates. An extreme case is the syndinians, where the distribution across the ten most abundant OTUS is nearly linear.

On the other extreme are lineages dominated by a single ‘jackpot’ OTU. This is the case for the diatoms, diplonemids, pelagophytes, and the dinoflagellates, where the jackpot is also the single most abundant OTU in the whole data set.

We might therefore ask, is the global OTU structure real, or it is an artifact of analytical pipelines and clustering methods? What are these jackpot taxa, and if they are real what does that tell us? Some of these questions have been addressed in different data sets by re-clustering the data using different methods [43], comparing the outcomes of different markers from similar environments [34], or, as noted above, using different levels of sampling from single cells to total communities [37].

A better sense for these limitations is important because tag sequencing guides our decisions as to where we focus our attention: without any other information about these organisms, abundance has become a proxy for importance, and without better methods to estimate abundance in a high-throughput manner, sequence abundance is the proxy for organismal abundance. And even if our estimates of abundance are acceptably close, its use as a proxy for importance is potentially misleading. We seldom have information about process rates from these organisms, so if, for example, rare organisms are more active than common ones abundance will not reflect ecological impact. Going back to the savanna, if you did what we do with microbes in the ocean, and scrapped up a patch of East Africa and counted the animals, you would find that there are lots of gazelles but just a few lions, but we know that lions have a huge impact on the systems. Interestingly, the results of our attempts to count marine microbial eukaryotes has not given what would perhaps be the most intuitive answer, that photosynthetic protists are more common: instead, heterotrophs dominate both at the

level of lineages and individual OTUs (Figure 3), emphasizing the need to understand what these counts really mean.

Remembering Cells in an Age of Genomes

In the absence of a wide diversity of model systems, environmental tag data will be interpreted as best we can, and functional data from potentially important but uncultivated species will be sought by other means. The same problem was faced by bacterial ecology years ago, but because of the different ways that eukaryotic diversity is manifested, environmental protistology will not be best served by simply following the trail blazed by bacteriology in the direction of bulk environmental sequencing, or meta-omics. Nevertheless, sequencing is relatively easy, can be informative, and is still the obvious first tool to apply to many questions. For many uncultivated protists, acquiring detailed biological observations can be nearly impossible, so it may be a question of acquiring the information that is available from sequence data, or nothing at all. There are also many downstream benefits of large sequence data sets even if they cannot be used to immediately infer detailed conclusions about the ecological role of the organism, as argued above. So, the question is not whether to sequence, but rather how to direct the tool at eukaryotic diversity so it is most informative, and not allowing our thinking to be restricted by treating sequencing as a panacea that will solve all our problems.

One point to consider is that cells really do matter. The ways that microbial eukaryotes rely on structure and behaviour to adapt are really only understood at the cellular level (as opposed to the whole community level), and even considering

metabolism alone, the partitioning of functions within and between cells is significant beyond the bulk metabolic function of an environment because of sub cellular complexity (e.g., organelles) and interactions like symbioses and phagotrophic feeding. Happily, we are in the midst of a technological breakthrough, where methods developed for high-throughput sequencing at the level of single-cells (primarily developed for medical research [44]) are maturing to the point of becoming widely applicable to uncultivated microbial species. Given the diversity of uncultivated protists, there is an almost limitless scope for sequencing: single-cell methods offer the chance retain a great deal of evolutionary context that allows us to overcome many of the technical difficulties that are inherent in analysing complex nuclear genomes.

Currently, single cell sequencing methods for protists include two broad approaches: single cell genomics (SCG, which generates data sets often referred to as single amplified genomes, or SAGs) [45], and single cell transcriptomics (SCT) [46]. As the names imply, SCG aims to characterize the whole genome at the DNA level, while SCT aims to characterize all the genes expressed as mRNA at the time the cell is collected. The benefits and drawbacks are similar to those of traditional genomics and transcriptomics. Acquiring a whole genome lends itself to a more comprehensive picture of the cell's potential, ideally including even those genes that are expressed at lower levels or under restricted conditions, as well as all the non-expressed sequences. But whole genomes are also more challenging to assemble (e.g., due to repeats) and interpret (e.g., without accurate gene models, introns and exons can be hard to predict). Also, current SCG technology does not generate complete genomes: in the few cases it has been applied to marine protists, an estimated 10–25% of the genome is reported,

which is only an estimate because the genome size itself is impossible to predict. In part, this is inevitable because only one or two copies of the genome exists in a single cell. SCG methods currently require an amplification step, which introduces biases and loss of information; so even sequencing to a great depth will not necessarily increase the coverage of the genome until amplification-independent sequencing methods are routinely brought to bear on single protist cells. Nevertheless, SCG-based studies of several marine protists have already led to new insights into mysterious and uncultivated marine protists that would not have been possible from bulk environmental sequencing. For example, the first description of the picobiliphyte protist as a new lineage of marine algae was based on microscopy and SSU rRNA from a single cell, leading to the conclusion that it **(Au: OK?)** was photosynthetic [47]. But a subsequent SCG-based analysis of three individual cells revealed they are not actually photosynthetic at all, and that the original observation was probably a food alga inside a heterotrophic predator [48]: picobiliphytes became picozoans [49]. Another recently discovered group are the marine diplonemids: long known to exist but considered a rare and ecologically uninteresting group, a subgroup of uncultured diplonemids was unexpectedly found to be among the most abundant marine heterotrophic eukaryotes in global marine surveys, as well as being incredibly diverse [36]. Using manually isolated cells and SCG analysis, their morphology was revealed along with a number of interesting features about genome structure and content that would not have been discernible from transcriptome data [50]. Similarly, SCG surveys of uncultivated Marine Alveoaltes (MALVs) (unpublished) and Marine Stramenopiles (MASTs) [51,52] have

given the first look at their genome content and the method has also been tested on the well-studied diatom *Thalassiosira* [51] and the parasite *Cryptosporidium parvum* [53].

In the history of genomic technology, transcriptomes (formerly called Expressed Sequence Tags, or ESTs) arrived relatively late as a means to quickly generate a large quantity of much more easily interpretable data specifically from expressed mRNA. The data are harder to generate, being derived from more finicky mRNA rather than from DNA, but because they are stripped of the introns and intergenic regions the genes are more readily assembled and analysed [27]. Transcriptomes are never a comprehensive survey of the genome or all its genes, only those that are expressed, and with SCT this means the subset of genes expressed at one time in a single cell, rather than an average across a population of cells in culture. The method also involves an amplification step, so bias and information loss also occur. Nevertheless, the first applications to protists revealed a promising rate of gene recovery: in a controlled study (of relatively large ciliates), SCT recovered between 80–100% of the genes recovered in comparable culture-derived transcriptomes from the same taxa, although with an apparent correlation between success and cell size [54]. It's unclear whether the biases introduced in SCT will be significant enough to challenge its application to questions relating to the response of gene expression levels to environmental change, or its use on extremely small eukaryotes [55], but for most uncultivated species at this point the goal is simply to acquire as much gene sequence as possible, and for this the method looks very promising.

Where it is possible to identify cells by morphology or some other criteria, the effectiveness of both SCT and SCG can be improved by pooling isolated cells or

pooling data derived from individual cells. This approach has been used to boost the quality of SCG assemblies [48,51,52] and to extract taxon-specific data from one member of a complex culture (e.g. a predator that feeds on a heterotrophic flagellate that itself feeds on bacteria [56]), but it could be used to overcome problems inherent in the lack of material that comes from dealing with single cells, and small cells in particular. At the extreme, pooling similar cells in very large numbers can be achieved by fluorescence activated cell sorting (FACS) technology. Given the appropriate sorting criteria, a relatively large population of cells can be extracted from a complex community using FACS. Ideally, the resulting population of cells will include only close relatives, but in practice several taxa with similar physical properties can be difficult to separate, resulting in sorted populations with more than one taxon. Even still, sorting can transform a complex community into a relatively simple one and sorted populations can be large enough to avoid the primary technical problems inherent in SCG and SCT [57–60]. Another potentially more powerful emerging method is microfluidics [61], which can separate complex communities into simplified populations of similar cells based on a wide range of micro reactions, which offers more and potentially more specific criteria to analyse individual species from complex protist communities. In marine ecological studies of bacteria, another common approach is metagenome assembled genomes (MAGs), where a genome, usually from an abundant organism, can be assembled from bulk environmental sequence data [62,63]. This has not been applied to marine microbial eukaryotes, and likely never will be very useful since there are few eukaryotic metagenomic data sets and it seems doubtful that nuclear genomes could be assembled in this way. One could more easily imagine metatranscriptome assembled

transcriptomes, but in this case the lack of physical linkage would make it difficult to bin transcripts to individual organisms, so again it seems doubtful such an approach would be more informative than a single cell method.

Conclusions

We live in an interesting time for the field of marine microbial ecology. It has been understudied for years, but the growing appreciation for the functional importance of marine microbial communities has coincided with the development of new genomic tools that have transformed it into a rapidly moving and changing field where major discoveries are taking place at a brisk pace.

But this does not mean we can simply sit back and reap the rewards. Indeed, we argue the opposite it true: for a deep understanding of marine microbial eukaryotes, now more than ever we need to examine where our tools are leading us and whether that is the right direction. It may be difficult but necessary to wean ourselves from the history of simply following the lead of environmental bacteriology. Sequences will never tell us everything, but for many microbial eukaryotes it may be currently the case that genomics won't tell us much at all about how they function in and respond to the environment because we lack detailed functional information from close relatives, which is necessary to make well-supposed inferences about homologous functions. Sequencing will still play a role so it is also important to ensure it's as effective as possible. On one hand, sequencing today is an investment in the future for when the more hard-won data from diverse model systems are available, since these data will

already be contextualized by sequences from many relatives. On the other hand, we lack a database of reference genomes from microbial eukaryotes that bacterial studies rely on to interpret environmental sequence data, and sequencing is necessary to fill this gap. This mean both a database of curated rRNAs to interpret sequence tags, but also complete genomes or transcriptomes to interpret environmental genomics such as metagenomics or metatranscriptomics. For nuclear genomes we seem to be trying to leapfrog the building of such a database and going straight to the generation of 'meta-omics' in the absence of anything to compare them to. We might learn more from greater emphasis on model system development, which would allow us to interpret environmental and also comparative data more effectively: comparative biology is one of the most powerful tools we have (in the absences of characterizing everything from everywhere), so improving the validity of our comparisons pays dividends. This will be particularly important when we start to look harder into the darker corners of the field — literally. Our knowledge of autotrophs is massively outstripping our understanding of heterotrophs, but heterotrophy is the ancestral state of eukaryotes, and a niche where they continue to play a huge role in the environment, and is perhaps even the most abundant form of microbial eukaryote (Figure 3). Heterotrophs are harder to study, but the potential rewards are even greater than the major advances made in algae.

Lastly, high-throughput molecular methods are fast and cheap, but we should not forget that culturing always will be an important part of understanding diversity, especially when it comes to some of the most interesting aspects of microbial eukaryotes such as their structure and behaviour. Culturing is time consuming, not always justly appreciated, and generally less fashionable than molecular methods, but

through cultures we discover functions and processes that have had disproportionately major impacts on how we see the role of microbial eukaryotes in the ocean [49,64–66].

References

1. Worden, A.Z., Follows, 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*, 1257594–1257594.
2. Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008). The microbial engines that drive earth's biogeochemical cycles. Science *320*, 1034–1039.
3. O'Malley, M.A. (2014). Philosophy of Microbiology 1st ed. (Cambridge, U.K.: Cambridge University Press).
4. Taylor, F.J.R. (1978). Problems in the development of an explicit hypothetical phylogeny of the lower eukaryotes. Biosystems *10*, 67–89.
5. del Campo, J., Sieracki, M.E., Molestina, R., Keeling, P., Massana, R., and Ruiz-Trillo, I. (2014). The others: our biased perspective of eukaryotic genomes. Trends Ecol. Evol. *29*, 252–259.
6. Rappé, M.S., and Giovannoni, S.J. (2003). The uncultured microbial majority. Annu. Rev. Microbiol. *57*, 369–394.
7. Massana, R. (2011). Eukaryotic picoplankton in surface oceans. Annu. Rev. Microbiol. *65*, 91–110.
8. Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., *et al.* (2016). A new

view of the tree of life. Nat. Microbiol. *1*, 16048.

9. Gilbert, J.A., and Dupont, C.L. (2011). Microbial metagenomics: beyond the genome. Ann. Rev. Mar. Sci. *3*, 347–371.
10. DeLong, E.F. (2009). The microbial ocean from genomes to biomes. Nature *459*, 200–206.
11. Schneider, T., and Riedel, K. (2010). Environmental proteomics: analysis of structure and function of microbial communities. Proteomics *10*, 785–798.
12. Caron, D.A., Worden, A.Z., Countway, P.D., Demir, E., and Heidelberg, K.B. (2009). Protists are microbes too: a perspective. ISME J. *3*, 4–12.
13. Lynch, M. (2003). The origins of genome complexity. Science *302*, 1401–1404.
14. Keeling, P.J., and Slamovits, C.H. (2005). Causes and effects of nuclear genome reduction. Curr. Opin. Genet. Dev. *15*, 601–608.
15. Keeling, P.J. (2013). Elephants in the room: protists and the importance of structure and behaviour. Environ. Microbiol. Rep. *5*, 4–5.
16. Madsen, E.L. (2011). Microorganisms and their roles in fundamental biogeochemical cycles. Curr. Opin. Biotechnol. *22*, 456–464.
17. Lee, J.J., Leedale, G.F., and Bradbury, P.C. eds. (2000). Illustrated Guide to the Protozoa 2nd ed. (Lawrence, Kansas, USA: The Society of Protozoologists).
18. Stolp, H., and Starr, M.P. (1963). Bdellovibrio bacteriovorus gen. et sp. n., a predatory, ectoparasitic, and bacteriolytic microorganism. Antonie Van Leeuwenhoek *29*, 217–248.
19. Keller, N.P., Turner, G., and Bennett, J.W. (2005). Fungal secondary metabolism — from biochemistry to genomics. Nat. Rev. Microbiol. *3*, 937–947.

20. Louca, S., Parfrey, L.W., and Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. Science *353*, 1272–1277.

21. Morris, R.M., Nunn, B.L., Frazar, C., Goodlett, D.R., Ting, Y.S., and Rocap, G. (2010). Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. ISME J. *4*, 673–685.

22. Saito, M.A., McIlvin, M.R., Moran, D.M., Goepfert, T.J., DiTullio, G.R., Post, A.F., and Lamborg, C.H. (2014). Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. Science *345*, 1173–1177.

23. Janouškovec, J., and Keeling, P.J. (2016). Evolution: causality and the origin of parasitism. Curr. Biol. *26*, R174–R177.

24. Zubkov, M.V., and Tarran, G.A. (2008). High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. Nature *455*, 224–226.

25. del Campo, J., and Parfrey, L.W. (2015). EukRef. The 18S collaborative annotation initiative. Available at: www.eukref.org [Accessed January 20, 2017].

26. Montagnes, D., Roberts, E., Lukeš, J., and Lowe, C. (2012). The rise of model protozoa. Trends Microbiol. *20*, 184–191.

27. Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., *et al.* (2014). the marine microbial eukaryote transcriptome sequencing project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. *12*, e1001889.

28. López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C., and Moreira, D. (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature

409, 603–607.

29. Moon-van der Staay, S.Y., de Wachter, R., and Vaulot, D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature *409*, 607–610.
30. Massana, R., and Pedrós-Alió, C. (2008). Unveiling new microbial eukaryotes in the surface ocean. Curr. Opin. Microbiol. *11*, 213–218.
31. del Campo, J., Guillou, L., Hehenberger, E., Logares, R., López-García, P., and Massana, R. (2016). Ecological and evolutionary significance of novel protist lineages. Eur. J. Protistol. *55*, 4–11.
32. Sogin, M.L., Morrison, H.G., Huber, J. a, Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., and Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere.” Proc. Natl. Acad. Sci. USA *103*, 12115–12120.
33. 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.
34. Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.-W., and Richards, T.A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol. Ecol. *19*, 21–31.
35. Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., Chambouvet, A., Christen, R., Claverie, J., Decelle, J., *et al.* (2015). Marine protist diversity in European coastal waters and sediments as revealed by high-throughput

sequencing. Environ. Microbiol. *17*, 4035–4049.

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

37. Heywood, J.L., Sieracki, M.E., Bellows, W., Poulton, N.J., and Stepanauskas, R. (2011). Capturing diversity of marine heterotrophic protists: one cell at a time. ISME J. *5*, 674–684.

38. Zhu, F., Massana, R., Not, F., Marie, D., and Vaulot, D. (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. FEMS Microbiol. Ecol. *52*, 79–92.

39. Rodríguez-Martínez, R., Labrenz, M., del Campo, J., Forn, I., Jürgens, K., and Massana, R. (2009). Distribution of the uncultured protist MAST-4 in the Indian Ocean, Drake Passage and Mediterranean Sea assessed by real-time quantitative PCR. Environ. Microbiol. *11*, 397–408.

40. Not, F., del Campo, J., Balagué, V., de Vargas, C., and Massana, R. (2009). New Insights into the Diversity of Marine Picoeukaryotes. PLoS One *4*, e7143.

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

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

43. Forster, D., Dunthorn, M., Stoeck, T., and Mahé, F. (2016). Comparison of three clustering approaches for detecting novel environmental microbial diversity. PeerJ

4, e1692.

44. Gawad, C., Koh, W., and Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. Nat. Rev. Genet. *17*, 175–188.

45. Stepanauskas, R. (2012). Single cell genomics: An individual look at microbes. Curr. Opin. Microbiol. *15*, 613–620.

46. Saliba, A.E., Westermann, A.J., Gorski, S.A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res. *42*, 8845–8860.

47. Not, F., Valentin, K., Romari, K., Lovejoy, C., Massana, R., Tobe, K., Vaulot, D., and Medlin, L.K. (2007). Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes. Science *315*, 253–255.

48. Yoon, H.S., Price, D.C., Stepanauskas, R., Rajah, V.D., Sieracki, M.E., Wilson, W.H., Yang, E.C., Duffy, S., and Bhattacharya, D. (2011). Single-cell genomics reveals organismal interactions in uncultivated marine protists. Science *332*, 714–717.

49. Seenivasan, R., Sausen, N., Medlin, L.K., and Melkonian, M. (2013). Picomonas judraskeda Gen. Et Sp. Nov.: The first identified member of the picozoa phylum Nov., a widespread group of picoeukaryotes, formerly known as “picobiliphytes.” PLoS One *8*, e59565.

50. Gawryluk, R.M.R., del Campo, J., Okamoto, N., Strassert, J.F.H., Lukeš, J., Richards, T.A., Worden, A.Z., Santoro, A.E., and Keeling, P.J. (2016). Morphological Identification and single-cell genomics of marine diplonemids. Curr. Biol. *26*, 3053–3059.

51. Roy, R.S., Price, D.C., Schliep, A., Cai, G., Korobeynikov, A., Yoon, H.S., Yang,

E.C., and Bhattacharya, D. (2014). Single cell genome analysis of an uncultured heterotrophic stramenopile. Sci. Rep. *4*, 4780.

52. Mangot, J., Logares, R., Sánchez, P., Latorre, F., Seeleuthner, Y., Mondy, S., Sieracki, M.E., Jaillon, O., Wincker, P., Vargas, C. de, *et al.* (2017). Accessing the genomic information of unculturable oceanic picoeukaryotes by combining multiple single cells. Sci. Rep. *7*, 41498.
53. Troell, K., Hallström, B., Divne, A., Alsmark, C., Arrighi, R., Huss, M., Beser, J., and Bertilsson, S. (2016). Cryptosporidium as a testbed for single cell genome characterization of unicellular eukaryotes. BMC Genomics *17*, 471.
54. Kolisko, M., Boscaro, V., Burki, F., Lynn, D.H., and Keeling, P.J. (2014). Single-cell transcriptomics for microbial eukaryotes. Curr. Biol. *24*, R1081–R1082.
55. Liu, Z., Hu, S.K., Campbell, V., Tatters, A.O., Heidelberg, K.B., and Caron, D.A. (2017). Single-cell transcriptomics of small microbial eukaryotes: limitations and potential. ISME J., 1–4.
56. Janouškovec, J., Tikhonenkov, D. V., Burki, F., Howe, A.T., Kolísko, M., Mylnikov, A.P., and Keeling, P.J. (2015). Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. Proc. Natl. Acad. Sci. USA *112*, 10200–10207.
57. Worden, A.Z., Janouškovec, J., Mcrose, D., Engman, A., Welsh, R.M., Malfatti, S., Tringe, S.G., and Keeling, P.J. (2012). Global distribution of a wild alga revealed by targeted metagenomics. Curr. Biol. *22*, R675–R677.
58. Lepère, C., Demura, M., Kawachi, M., Romac, S., Probert, I., and Vaulot, D. (2011). Whole-genome amplification (WGA) of marine photosynthetic eukaryote

populations. FEMS Microbiol. Ecol. *76*, 513–523.

59. Cuvelier, M.L., Allen, A.E., Monier, A., McCrow, J.P., Messie, M., Tringe, S.G., Woyke, T., Welsh, R.M., Ishoey, T., Lee, J.-H., *et al.* (2010). Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. Proc. Natl. Acad. Sci. USA *107*, 14679–14684.

60. Martinez-Garcia, M., Brazel, D., Poulton, N.J., Swan, B.K., Gomez, M.L., Masland, D., Sieracki, M.E., and Stepanauskas, R. (2012). Unveiling in situ interactions between marine protists and bacteria through single cell sequencing. ISME J. *6*, 703–707.

61. Rusconi, R., Garren, M., and Stocker, R. (2014). Microfluidics expanding the frontiers of microbial ecology. Annu. Rev. Biophys. *43*, 65–91.

62. Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., and Nielsen, P.H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat. Biotechnol. *31*, 533–538.

63. Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., and Ettema, T.J.G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. Nature *521*, 173–179.

64. Moore, R.B., Oborník, M., Janouškovec, J., Chrudimský, T., Vancová, M., Green, D.H., Wright, S.W., Davies, N.W., Bolch, C.J.S., Heimann, K., *et al.* (2008). A photosynthetic alveolate closely related to apicomplexan parasites. Nature *451*, 959–963.

65. del Campo, J., Not, F., Forn, I., Sieracki, M.E., and Massana, R. (2013). Taming the smallest predators of the oceans. ISME J. *7*, 351–358.

66. Shalchian-Tabrizi, K., Eikrem, W., Klaveness, D., Vaulot, D., Minge, M.A., Le Gall, F., Romari, K., Throndsen, J., Botnen, A., Massana, R., *et al.* (2006). Telonemia, a new protist phylum with affinity to chromist lineages. Proc. R. Soc. B Biol. Sci. *273*, 1833–1842.

67. Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Verezemska, O., Isbandi, M., Thomas, A.D., Ali, R., Sharma, K., Kyrpides, N.C., *et al.* (2017). Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. Nucleic Acids Res. *45*, D446–D456.

68. Leinonen, R., Sugawara, H., and Shumway, M. (2011). The sequence read archive. Nucleic Acids Res. *39*, 2010–2012.

69. Mitchell, A., Bucchini, F., Cochrane, G.R., Denise, H., Ten Hoopen, P., Fraser, M., Pesseat, S., Potter, S., Scheremetjew, M., Sterk, P., *et al.* (2016). EBI metagenomics in 2016 - An expanding and evolving resource for the analysis and archiving of metagenomic data. Nucleic Acids Res. *44*, D595–D603.

70. Hurwitz, B.L., Youens-Clark, K., and Walls, R.L. (2017). iMicrobe. Available at: http://imicrobe.us/ [Accessed January 20, 2017].

Figure 1. Examples of morphological and trophic complexity of marine microbial eukaryotes**.**

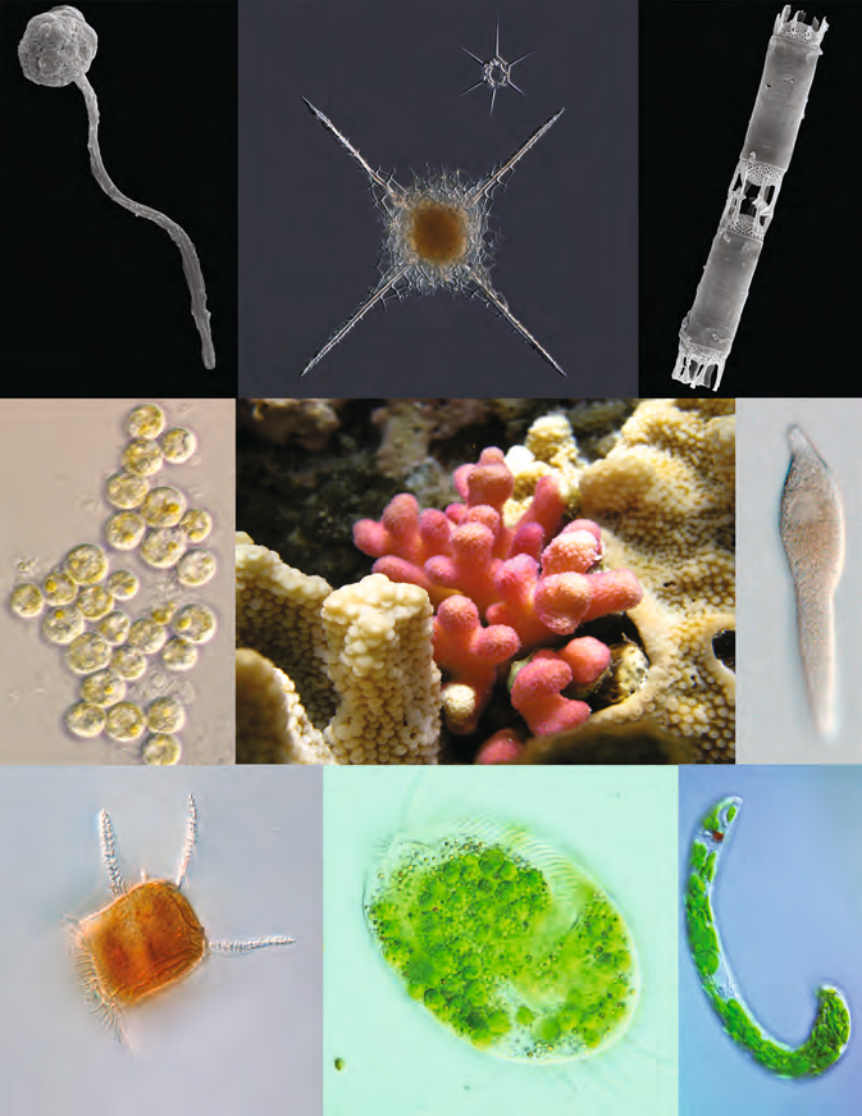
Top row (left to right): the heterotrophic rhizarian nanoflagellate *Minorisa*, which is

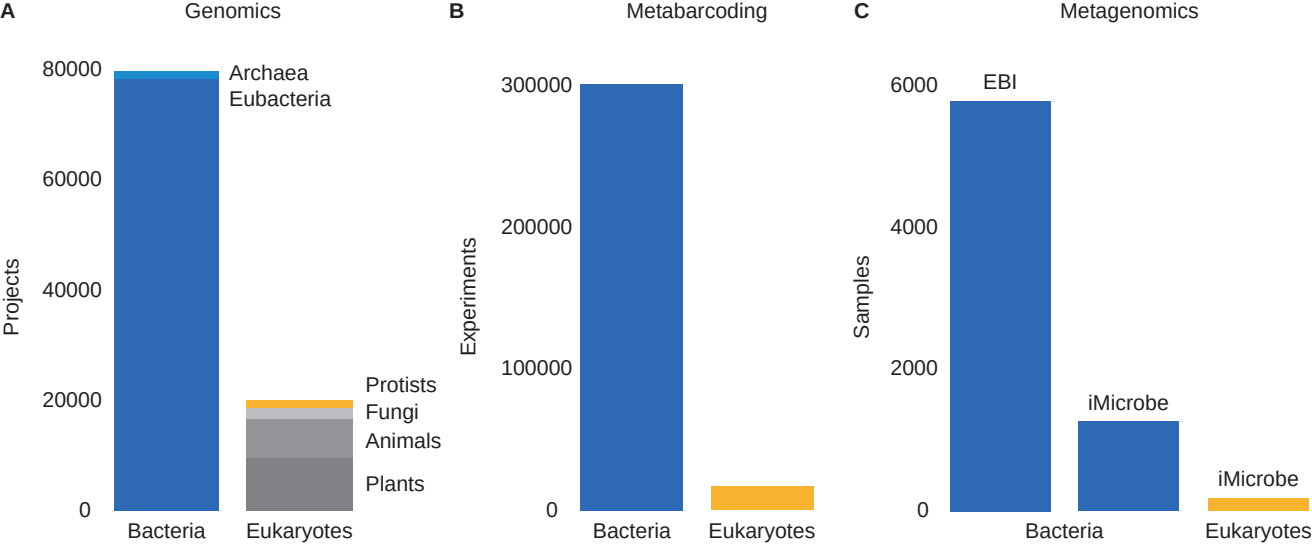
common in coastal waters; a large heterotrophic radiolarian, *Rhizoplegma*, which are among **(Au: Something seems to be missing here)**; a diatom, one of the most common autotrophic eukaryotes. Centre row (left to right): the dinoflagellate *Symbiodinum*, a photosynthetic endosymbiont of corals (centre), which give corals their distinctive colours and make reef-building possible; the gregarine apicomplexan, *Lankesteria*, which are well-studied parasites of terrestrial animals (like us), but are also abundant in the ocean and likely play a significant role in marine animal populations. Bottom row (left to right): the photosynthetic dinoflagellate *Ceratocorys*, with its distinctive cellulose spines; the ciliate *Euplotes*, which is a predator but here appears green because it is filled with prey algae (*Duniella*); the photosynthetic euglenozoan *Euglena*, which is also green but here due to its green secondary plastids. It also has a visible red 'eyespot' which it uses to locate light. All photos by the authors except: *Rhizoplegma* (© John Dolan), *Symbiodinum* and *Ceratocorys* (courtesy of Nick Irwin), and *Lankesteria* (© Sonja Rueckert).

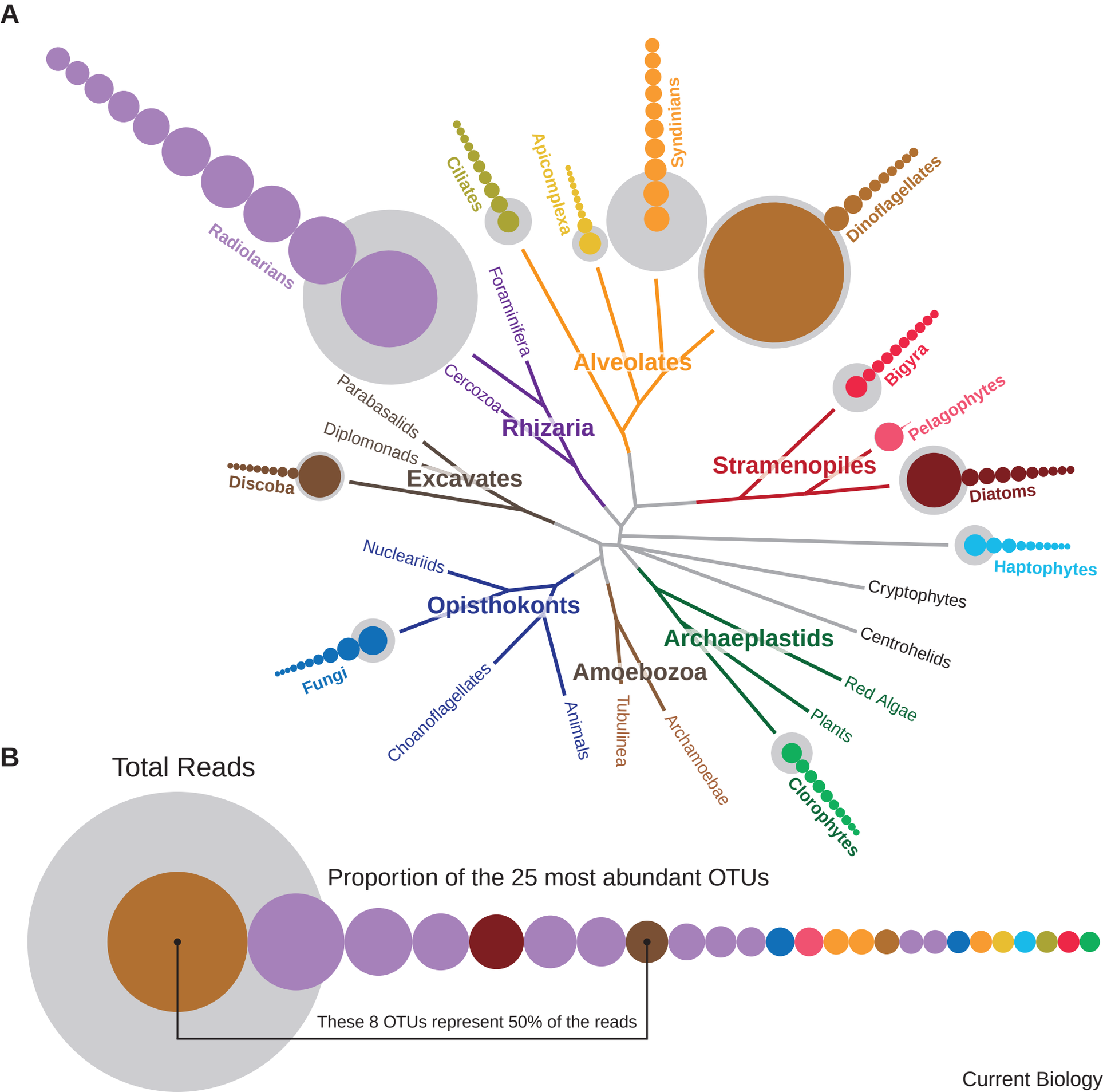
Figure 2. Current microbial genomic data is heavily biased towards bacteria.
(A) The current number of completed genomic projects (retrieved from The Genomes OnLine Database [67]). (B) The current number of tag sequencing studies using 16S rRNA for bacteria as opposed to 18S rRNA for eukaryotes (retrieved from the Sequence Read Archive [68]). (C) The current number of environmental metagenomics and metatranscriptomic studies in EBI Metagenomes database [69] and iMicrobe database [70]. We looked at two databases because EBI Metagenomes, although the largest, contains no samples identified as eukaryotic.

Figure 3. Relative abundance patterns of the most common microbial eukaryotic OTUs in Tara Oceans.

(A) The relative abundance of the most common operational taxonomic units (OTUs) within 12 well-defined and diverse protist lineages. For each lineage, the total number of reads for the entire group is shown as a grey circle (the size shown to scale between lineages), and the ten most common individual OTUs (e.g., species) are shown as coloured circles of descending size. (B) The 25 most abundant protist OTUs in the entire Tara Oceans data set [41]. As above, the grey circle represents the size of the whole data set, while the coloured circles represent individual OTUs, colour coded according to lineage as in panel A. The first eight OTUs account for over 50% of the total number of reads in the entire data set.




A
Genomics
Projects
80000
60000
40000
20000
0
Archaea
Eubacteria
Protists
Fungi
Animals
Plants
Bacteria
Eukaryotes
B
Metabarcoding
Experiments
300000
200000
100000
0
Bacteria
Eukaryotes
C
Metagenomics
Samples
6000
4000
2000
0
EBI
iMicrobe
iMicrobe
Bacteria
Eukaryotes


A
Radiolarians
Ciliates
Apicomplexa
Syndinians
Dinoflagellates
Foraminifera
Alveolates
Cercozoa
Rhizaria
Parabasalids
Diplomonads
Excavates
Discoba
Bigyra
Pelagophytes
Stramenopiles
Diatoms
Haptophytes
Cryptophytes
Centrohelids
Nucleariids
Opisthokonts
Fungi
Archaeplastids
Red Algae
Plants
Clorophytes
Choanoflagellates
Animals
Amoebozoa
Tubulinea
Archamoebae
B
Total Reads
Proportion of the 25 most abundant OTUs
These 8 OTUs represent 50% of the reads
Current Biology