

## Probing the evolution, ecology and physiology of marine protists using transcriptomics

David A. Caron<sup>1</sup>, Harriet Alexander<sup>2</sup>, Andrew E. Allen<sup>3,4</sup>, John M. Archibald<sup>5,6</sup>, E. Virginia Armbrust<sup>7</sup>, Charles Bachy<sup>8</sup>, Callum J. Bell<sup>9</sup>, Arvind Bharti<sup>10</sup>, Sonya T. Dyhrman<sup>11</sup>, Stephanie M. Guida<sup>9</sup>, Karla B. Heidelberg<sup>1</sup>, Jonathan Z. Kaye<sup>12</sup>, Julia Metzner<sup>12</sup>, Sarah R. Smith<sup>4</sup> and Alexandra Z. Worden<sup>6,8</sup>

**Abstract** | Protists, which are single-celled eukaryotes, critically influence the ecology and chemistry of marine ecosystems, but genome-based studies of these organisms have lagged behind those of other microorganisms. However, recent transcriptomic studies of cultured species, complemented by meta-omics analyses of natural communities, have increased the amount of genetic information available for poorly represented branches on the tree of eukaryotic life. This information is providing insights into the adaptations and interactions between protists and other microorganisms and macroorganisms, but many of the genes sequenced show no similarity to sequences currently available in public databases. A better understanding of these newly discovered genes will lead to a deeper appreciation of the functional diversity and metabolic processes in the ocean. In this Review, we summarize recent developments in our understanding of the ecology, physiology and evolution of protists, derived from transcriptomic studies of cultured strains and natural communities, and discuss how these novel large-scale genetic datasets will be used in the future.

### Phototrophy

A nutritional mode that involves the use of light for the production of organic carbon and the acquisition of energy.

### Phytoplankton

Planktonic protists that use phototrophy as their nutritional mode. The term has ecological importance but little phylogenetic importance because the behaviour occurs across many lineages of protists.

Correspondence to D.A.C.  
Department of Biological Sciences, University of Southern California, 3616 Trousdale Parkway, Los Angeles, California 90089, USA.  
[dcaron@usc.edu](mailto:dcaron@usc.edu)

doi:10.1038/nrmicro.2016.160  
Published online 21 Nov 2016

Marine protists are a phylogenetically broad group of eukaryotic organisms that are capable of existence as single cells, although many form colonies that exhibit coordinated behaviour. The term protist is largely historical in nature, at one time referring to a kingdom-level distinction in Whittaker's tree of life<sup>1</sup>, but still commonly applied to all eukaryotes that are not plants, multicellular animals or fungi. Most protists are microscopic, but collectively these species span more than five orders of magnitude in size, display a myriad of morphologies, behaviours and nutritional modes, and have pivotal ecological roles in marine ecosystems<sup>2,3</sup>. Protistan species are the most abundant eukaryotes in marine pelagic communities and constitute a substantial portion of the total biomass in the plankton.

Nutritional modes of protists range from pure phototrophy among many phytoplankton to the pure heterotrophy of species traditionally called protozoa. Protozoa generally derive their nutrition from the ingestion of bacteria, archaea or other eukaryotes, but some are parasites (or more accurately, parasitoids) of other protists and metazoa. In addition, many protists combine phototrophy and heterotrophy in a single species (mixotrophy), such as phytoflagellates that have

chloroplasts but also consume prey. Protists typically reproduce through mitotic division rather than through sexual reproduction, which enables populations to double in a few hours to a few days. Their rapid growth rates enable them to contribute to important ecosystem functions: protists have roles in primary production, as partners in various symbiotic relationships and, at multiple trophic levels, as links between the small animals that prey on them and the vast numbers of bacteria, archaea, protists and even some metazoa that they consume (FIG. 1). These links in the food web are essential for the biogeochemical cycles of the ocean, including the transport of carbon to the deep ocean.

Despite the fundamental roles of protists in marine ecosystems, genome-based studies of these species have lagged behind those of other marine organisms. However, in the past few decades, a lot of DNA sequence information has become available for protistan marker genes<sup>4</sup>, transcriptomes<sup>5</sup>, genomes (almost 20 complete genomes in total)<sup>6</sup>, metatranscriptomes<sup>5</sup> and even metagenomes<sup>7</sup>. Analyses of these data have altered our understanding of protistan diversity, phylogeny, evolution and physiology, and enriched our knowledge of various ecological interactions.

## Author addresses

<sup>1</sup>Department of Biological Sciences, University of Southern California, 3616 Trousdale Parkway, Los Angeles, California 90089, USA.

<sup>2</sup>Department of Population Health and Reproduction, School of Veterinary Medicine, 1 Shields Avenue, University of California Davis, Davis, California 95616, USA.

<sup>3</sup>Microbial and Environmental Genomics, J. Craig Venter Institute, 4120 Capricorn Lane, La Jolla, San Diego, California 92037, USA.

<sup>4</sup>Scripps Institution of Oceanography, Integrative Oceanography Division, University of California San Diego, 9500 Gilman Drive, La Jolla, San Diego, California 92093, USA.

<sup>5</sup>Department of Biochemistry and Molecular Biology, Dalhousie University, Sir Charles Tupper Medical Building, 5850 College Street, Halifax, Nova Scotia B3H 4R2, Canada.

<sup>6</sup>Canadian Institute for Advanced Research, 180 Dundas Street W, Toronto, Ontario M5G 1Z8, Canada.

<sup>7</sup>School of Oceanography, University of Washington, 1503 NE Boat Street, Seattle, Washington 98195, USA.

<sup>8</sup>Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, California 95039, USA.

<sup>9</sup>National Center for Genome Resources, 2935 Rodeo Park Drive East, Santa Fe, New Mexico 87505, USA.

<sup>10</sup>Syngenta Biotechnology, 629 Davis Drive, Research Triangle Park, North Carolina 27709, USA.

<sup>11</sup>Department of Earth and Environmental Sciences and the Lamont-Doherty Earth Observatory, Columbia University, 61 Route 9W, Palisades, New York 10964, USA.

<sup>12</sup>Gordon and Betty Moore Foundation, 1661 Page Mill Road, Palo Alto, California 94304, USA.

### Heterotrophy

A nutritional mode that involves the use of preformed organic matter for the acquisition of carbon and energy.

### Protozoa

Protists that are not photosynthetic, but are instead dependent on the ingestion of preformed organic matter (usually prey) for their nutrition. This older term is still in use; 'heterotrophic protists' is used synonymously.

### Parasitoids

Protists that exhibit a parasitic lifestyle, infecting other protists or multicellular organisms. Parasitoids are typically differentiated from parasites by the fact that parasitoids, unlike parasites, always kill their hosts.

### Metazoa

Multicellular, eukaryotic organisms (animals) that have differentiated cells and tissues.

### Mixotrophy

A nutritional mode in which an individual cell can use both inorganic and preformed organic sources of carbon and nutrients for growth. In protists, mixotrophy is generally accomplished by combining heterotrophy with a photosynthetic ability acquired through chloroplasts, kleptoplastidy or harbouring endosymbiotic algae.

The historical partitioning of eukaryotes into kingdoms based on morphological characteristics, multicellularity or trophic mode<sup>1</sup> has been challenged and changed in recent years by information on genetic relatedness. The resulting framework for a molecular taxonomy for protists has raised questions regarding the biogeography of these taxa and the adequacy of the classical morphological species concept for protists<sup>8–10</sup>. Direct sequencing of rRNA genes from environmental samples has revealed a huge protistan diversity, including new lineages<sup>11,12</sup>, and has stimulated hypotheses regarding the protistan rare biosphere — diverse taxa present at low relative abundances in virtually all natural ecosystems — which may play important parts in the evolution of eukaryotes as well as in the stability and functional resilience of microbial communities<sup>13–15</sup>. Importantly, studies of protistan gene expression have begun to provide unprecedented insight into how protists respond to environmental cues, competition, predation and symbiotic interactions with other microorganisms.

In this Review, we summarize recent developments in the characterization of protistan gene expression, discuss how transcriptomic studies are revealing novel insights into protistan biology, and explain how these findings will influence our understanding of ocean ecology and biogeochemistry.

### Filling in the blanks

Compared with other microorganisms, microbial eukaryotes have been the 'poor cousins' of the omics revolution<sup>6</sup>, in large part owing to their larger and more complex genomes<sup>16</sup>. The eukaryotic genomes that have been sequenced to date are dominated by metazoan model organisms, a handful of microorganisms that parasitize humans (for example, *Plasmodium* spp., *Trypanosoma* spp. and *Giardia lamblia*) and a few photosynthetic marine microbial eukaryotes. Consequently, DNA sequence databases that are devoted to free-living, ecologically important protists remain sparse.

A few large culture-based and field-based projects have begun to address this disparity. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) released nearly 680 transcriptome assemblies of pure cultures of protists in 2014, representing isolates from much of the world's oceans and a few freshwater ecosystems (BOX 1). The MMETSP was conceived to facilitate comparative transcriptomics of environmentally important microbial eukaryotes. The dataset increased ~50-fold the amount of information available on gene expression from free-living protists, and doubled the number of protistan lineages for which genomic studies have been conducted<sup>5</sup>. Subsequent work has continued to improve the gene assemblies<sup>17</sup>. Concomitantly, two field programmes (the Tara Oceans and Malaspina expeditions)<sup>4,18</sup> have circumnavigated the planet collecting samples to provide an enormous sample set for the investigation of the abundances, morphologies, genetic diversity and ecological significance of marine protists. These studies highlight the importance of marine protists, demonstrate the genetic and physiological potentials that underpin their activities and provide new insights into the evolution of these species.

### Insights into protistan phylogeny and evolution

**Modern protistan phylogeny.** Our view of the evolutionary relationships among protistan lineages has changed substantially during the past few decades (FIG. 2). The Whittaker system<sup>1</sup>, which prevailed for two decades, grouped organisms into five kingdoms: Monera, Protista, Plantae, Fungi and Animalia. Eukaryotes that can exist as single cells were contained within the kingdom Protista. Protists were further organized according to morphology and nutritional preferences, a scheme that was practical for biologists of the time but artificially separated some closely related taxa that have different nutritional behaviours (for example, photosynthetic and heterotrophic dinoflagellates).

The adoption of ultrastructural and DNA sequence information eventually undermined Whittaker's scheme<sup>19</sup>. The 'new' system derived from this information placed all eukaryotes into a single domain, the Eukaryota, with the animals, plants and fungi represented as minor branches among a huge diversity of protists<sup>20</sup>. Nearly 15 years of reorganization followed, during which all eukaryotes were arranged into 'super-groups', a scheme that is believed to more accurately reflect evolutionary history. The scheme continues to undergo numerous minor revisions to resolve the root and branching order of various protistan groups (FIG. 2).

The result is a domain of eukaryotic life, the domain Eukarya, that bears little resemblance to Whittaker's scheme at the broadest phylogenetic groupings, although many of the protistan phyla remained intact throughout the reorganization<sup>21,22</sup> (FIG. 2). Reorganization within and among some of the supergroups, such as the division of amoeboid protists into the Amoebozoa and Rhizaria<sup>23</sup>, and further grouping of some of the supergroups into higher order 'clusters' (for example, the grouping of the Stramenopiles, Alveolata and Rhizaria supergroups into the SAR cluster)<sup>24</sup> continue to refine the structure of the

**Primary production**

The photosynthetic production of organic carbon, carried out by a wide variety of protists, macroalgae and plants.

**Metatranscriptomes**

Collections of all the transcriptomes (all RNA transcripts) present in communities of microorganisms; a metatranscriptome of a community is derived from RNA extraction and purification, reverse transcription of RNA to cDNA and sequencing of the resulting cDNA.

**Metagenomes**

Collections of all the DNA present in communities of microorganisms, representing all the genetic potential of the communities. The metagenome of a community can be used to reconstruct the genomes of the individual species comprising that community, thus assigning specific metabolic roles to those taxa.

**Dinoflagellates**

Members of a major, flagellated protistan lineage (the class Dinophyta, in the supergroup Alveolata) containing phototrophic, heterotrophic and mixotrophic (kleptoplastidic) species. Numerous photosynthetic and mixotrophic dinoflagellates are harmful and produce toxic algal blooms that have traditionally been called red tides.

**Amoebozoa**

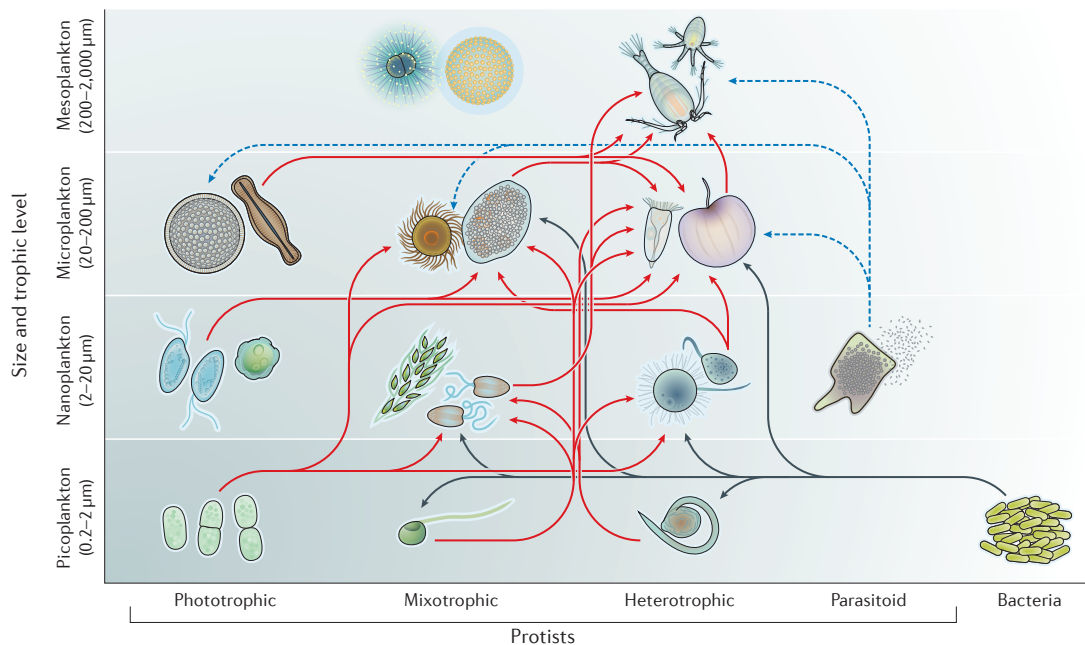
A protist supergroup that includes many small amoeboid forms and the slime moulds.

**Rhizaria**

A large protist supergroup that includes ecologically important, large amoeboid forms (radiolaria and foraminifera) and the Cercozoa.

**Stramenopiles**

A diverse protist supergroup of phototrophic, heterotrophic and mixotrophic (phagotrophic phytoflagellates) species characterized by the presence of two flagella of unequal length and structure in their motile life stages. The supergroup includes the brown algae, the chrysophytes, the diatoms and other important groups. The term is generally used synonymously with heterokonts.



**Figure 1 | Ecological and biogeochemical roles of protists in the marine plankton.** Protists are an important part of living biomass in oceanic ecosystems and are central to a wide array of food web processes and biogeochemical cycles. Species of phototrophic, heterotrophic and mixotrophic protists span more than three orders of magnitude in size and several trophic levels at, and near, the base of the food web. Complex predator–prey relationships that involve protists make photosynthetically produced organic material (red arrows) and bacterial biomass (black arrows) available to higher trophic levels and also remineralize a substantial fraction of this material back to inorganic nutrients and carbon dioxide to support new primary production. Recent environmental gene surveys suggest that protistan parasitoids are also important players in marine food webs, as they prey on microplanktonic and mesoplanktonic species<sup>112</sup> (blue dashed arrows). In addition to the myriad of predator–prey relationships, the interactions of protists with other microorganisms and multicellular organisms include a wide range that are not depicted here (involving competition, commensalism and mutualism).

domain, whereas relationships among other protistan groups remain unresolved at this time<sup>25,26</sup>. These uncertainties notwithstanding, the basic topology within the eukaryotic tree has finally begun to stabilize<sup>26</sup>.

rRNA genes are slow to evolve and so have been used extensively to redefine the high-level phylogenetic relationships among eukaryotes. Studies are continuing, but now with the use of multi-gene datasets, which are necessary to increase resolution; much of this work relies heavily on transcriptomic datasets because of the present scarcity of sequenced genomes for microbial eukaryotes. Grant and Katz<sup>27</sup>, for example, used transcriptomic data that encompassed >15,000 characters (including concatenated sequences of ribosomal genes and >200 protein-coding genes) to establish the phylogenetic placement of several species of amoebae within the Amoebozoa.

**Symbiosis, endosymbiosis and organelle acquisition.**

The evolution of mitochondria and chloroplasts from bacteria through endosymbiosis were pivotal events in the history of the eukaryotic cell. Present-day eukaryotes are thought to have diversified from a mitochondrion-bearing ancestor<sup>28,29</sup>, and chloroplasts evolved from cyanobacterial endosymbionts in a common ancestor that is shared by red algae, green algae (and their land plant relatives) and glaucophytes<sup>30</sup>. Secondary and

tertiary endosymbiotic events then spread the so-called primary chloroplasts of red and green algae to an array of organisms<sup>31,32</sup>, including euglenids, diatoms, haptophytes and dinoflagellates. The construction of ‘phylogenomic’ frameworks for understanding the relationships between photosynthetic and heterotrophic lineages, and how chloroplasts have been gained and lost, has been instrumental for documenting gene transfer events between symbiont and host<sup>33</sup>. Transcriptome-based studies have played a key part in this process. For example, gene discovery surveys have revealed the presence of novel chloroplast genes in heterotrophic organisms<sup>34</sup>.

Algae that have secondary chloroplasts are among the most genetically and biologically complex cells known. The cryptophyte *Guillardia theta* and the chlorarachniophyte *Bigelowiella natans* are two unrelated algal species that both have secondary chloroplasts<sup>31</sup>, but they are unusual in that the nucleus of the secondary endosymbiont persists in the form of a vestigial ‘nucleomorph’ (REF. 35). The nuclear genomes of both organisms have now been sequenced, and the data supplemented with transcriptomic sequence information<sup>36</sup>. Both organisms are predicted to have >21,000 intron-rich nuclear genes. Genes from *B. natans* showed a level of alternative intron splicing that exceeded that seen in genes from more complex organisms, such as the model plant *Arabidopsis thaliana*, and on par with the level seen

Box 1 | The Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP)

The Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP) was a global community effort to augment the available DNA sequence information on ecologically relevant protists and provide a much needed reference dataset for environmental metagenomic and metatranscriptomic studies. The overall goal was to substantially increase the sequence datasets from cultured, well-curated protistan species to provide a resource for gene annotation throughout the Eukarya and for the interpretation of metagenomic and metatranscriptomic datasets from diverse marine ecosystems. Supported by the Gordon and Betty Moore Foundation, the project directly involved nearly 70 laboratories engaged in protistan research, and more than 200 investigators worldwide participated by submitting material for sequencing. The project entailed the sequencing and assembly of 678 transcriptomes that encompassed 210 unique genera, 305 species and 396 strains of protists, and the results were released to the public in June 2014. Depicted in the figure are the geographical origins of approximately half of the strains, indicating the degree of global coverage. The MMETSP sequence datasets are publicly available through the [iMicrobe](#) and [NCBI BioProject](#) websites, and recently re-annotated datasets are available at the [MMETSP re-assemblies](#) website. A list of publications that resulted from the project is available at the [Gordon and Betty Moore Foundation MMETSP webpage](#). An interactive version of the map indicating the locations from which strains were obtained is available on the [iMicrobe](#) site. A list of MMETSP participants and their affiliations are provided ([Supplementary information S1](#) (box)).



**Alveolata**

A supergroup that includes three important groups of protists: the dinoflagellates, the ciliates and the parasitic apicomplexans. The defining characteristic of the Alveolata is the presence of alveoli (flattened vesicles beneath the cell wall).

**Glaucophytes**

(Also known as glaucocystophytes). Members of the class Glaucocystophyceae, a small algal clade that is grouped together with the green algae and land plants in the supergroup Archaeplastida.

**Euglenids**

Members of the phylum Euglenida (in the supergroup Excavata). Free-living euglenids include phototrophic, heterotrophic and mixotrophic species, and a few notable parasites of animals.

**Diatoms**

Members of the phylum Bacillariophyta (in the supergroup Stramenopiles), a clade that is characterized by siliceous cell walls.

**Haptophytes**

Members of an algal group that encompasses prymnesiophytes (supergroup unresolved), including the bloom-forming species of the genus *Phaeocystis* as well as globally and biogeochemically important forms such as the coccolithophorids, which often bear calcium carbonate plates (coccoliths).

**Cryptophyte**

A member of the class Cryptophyta (supergroup unresolved); this is a group of small, flagellated protists that contains mostly photosynthetic forms but also mixotrophic and heterotrophic species. The photosynthetic species have plastids that contain phycobiliprotein pigments and are derived from red algae.

in genes expressed in the human cerebral cortex<sup>36</sup>. The role of such rampant alternative splicing is unclear and is being investigated.

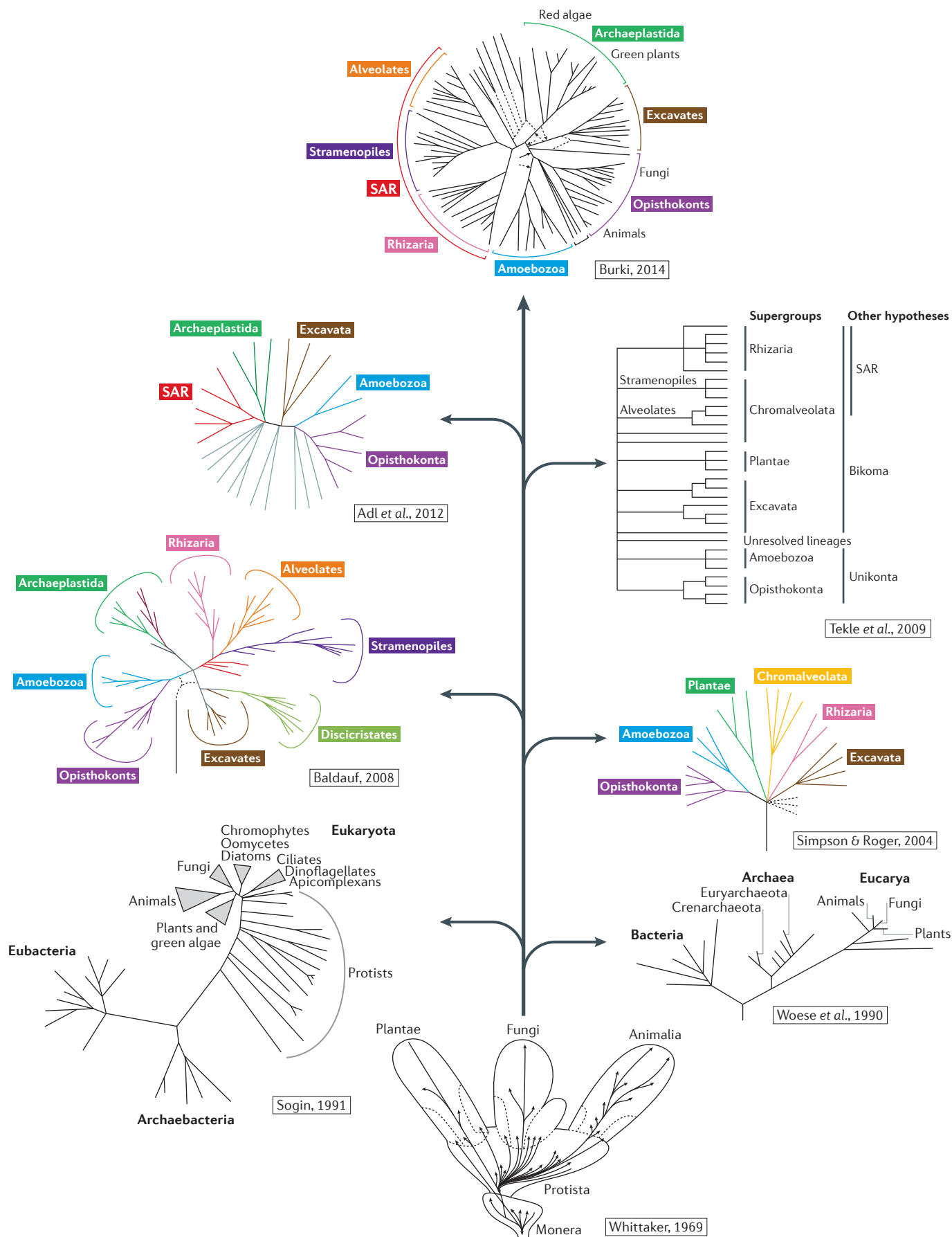
One particularly exciting example of ‘evolution in action’ is the enigmatic rhizarian *Paulinella chromatophora*, for which transcriptomic analyses have revealed that at least three genes in the nuclear genome encode subunits of photosystem I in the chromatophore — the recently established endosymbiont of *P. chromatophora*<sup>37,38</sup>. Transcripts from these genes are translated in the host cytoplasm and the resultant proteins are targeted back to the chromatophore<sup>39</sup>. The chromatophore thus qualifies as a cyanobacterium-derived photosynthetic organelle of recent origin that has evolved completely independently of classical chloroplasts<sup>40</sup>. Transcriptomic studies of such transitional states of organelle acquisition are providing a greater understanding of the underlying interactions between endosymbionts and their hosts<sup>41</sup>.

A myriad of endosymbiotic associations, such as those that involve large rhizarians and photosynthetic protists (FIG. 3a), evolved more recently than the origin of mitochondria and chloroplasts. Understanding how these associations are established and maintained is

fundamental for understanding the evolution of complex life<sup>41</sup>. Coral–dinoflagellate (*Symbiodinium* spp.) symbioses have become a model system for studying metazoan–protist symbiosis. Transcriptomic approaches (and the recent sequencing of the *Symbiodinium kawagutii* nuclear genome<sup>42</sup>) have become fundamental tools for unravelling the role of the dinoflagellates in these associations<sup>43</sup>. Approximately half of the dinoflagellate genes do not match genes from any other organism, and a small percentage of these genes are regulated transcriptionally. Moreover, biochemical complementarity between symbiont and host seems to indicate host–symbiont co-evolution, implying that the interaction involves an energetic cost efficiency, which may explain the ecological success of the association.

Progress towards understanding the associations between rhizarians and photosynthetic protists, which are exceptionally abundant in the ocean plankton<sup>44,45</sup>, has been slow because of challenges in culturing most host species. Culture-independent transcriptomic approaches provide a way around this obstacle. For example, the comparison of transcriptomes from four rhizarians (two polycystines, an acantharian and a phaeodarian) revealed the presence of genes that encode c-type lectins





## Chlorarachniophyte

A member of a group of algae (Chlorarachniophyceae) that are often mixotrophic. These species are also often amoeboid in form and are placed in the supergroup Rhizaria within the Cercozoa, together with large amoeboid forms, such as radiolaria and foraminifera.

## Nucleomorph

A vestigial nucleus that is associated with plastids in some protists. They are derived from the engulfment and reduction of a eukaryotic endosymbiont.

## Chromatophore

(In this Review:) An endosymbiotic cyanobacterium with a reduced genome. In the protist *Paulinella chromatophora*, this structure is considered an early stage of chloroplast acquisition.

## Plastid

One of several types of cyanobacterium-derived, double-membrane-bound organelles that are present in many protists; an example is the chloroplast.

in the polycystines and acantharian, which harbour dinoflagellate endosymbionts, but not in the phaeodarian, which is symbiont-free<sup>46</sup>. These proteins may play a part in establishing the endosymbiosis because they are known to be involved in cell–cell interactions in other species. Other transcriptomic studies have been carried out on rhizarians that retain the functional chloroplasts of ingested diatom prey; these studies have investigated how plastid function is maintained and the possibility of horizontal gene transfer between host and prey<sup>47</sup>.

**Algal phytochromes.** Transcriptomics has also transformed our view of evolution within and among algal groups that have sequenced genomes, such as green algae (also known as chlorophytes) and glaucophytes. A recent example is the discovery of widespread sensory proteins that react to changes in light. In plants, phytochromes are involved in numerous processes, including seed germination, stem elongation and reproduction<sup>48</sup>. Until recently, these organelles were thought to have limited distributions among other eukaryotes, being mainly restricted to fungi and stramenopiles (in particular, diatoms)<sup>49–51</sup>; they were found in a single prasinophyte genome (*Micromonas pusilla*)<sup>52</sup> but were absent from the other prasinophyte genomes studied and also from the genomes of other studied green algae, such as the model alga *Chlamydomonas reinhardtii*. However, increased transcriptomic coverage of environmental lineages has revealed phytochromes in the majority of the prasinophyte groups (that is, in species from six of the seven accepted clades), indicating that they have a broader distribution in green algae than had been realized previously. Transcriptomic data also revealed their presence in glaucophytes (which are in

the same eukaryotic supergroup as green algae and land plants) and in algae from other supergroups, specifically in other stramenopiles (beyond diatoms) and in cryptophytes<sup>53</sup>. Some of these new findings may simply be the result of expanding the datasets from taxa with larger genomes, given the focus of early genome sequencing projects on widespread taxa with small genomes (in which gene loss is often substantial)<sup>54</sup>.

## Insights into protistan physiology and ecology

Transcriptomic data are now available for a greatly expanded array of protistan taxa, and this has facilitated investigations of gene expression in protists, in particular for many important primary producers, and, to a lesser extent, for a few heterotrophic and mixotrophic species. Such studies are beginning to elucidate the specific physiological responses of these protists to environmental cues and are identifying the genes that are associated with specific nutritional strategies. New gene discovery within various microalgal lineages in particular has moved forwards rapidly, resulting in a substantial increase in the number of known protist protein families and a clearer understanding of their evolutionary origins and metabolic functions<sup>55</sup>.

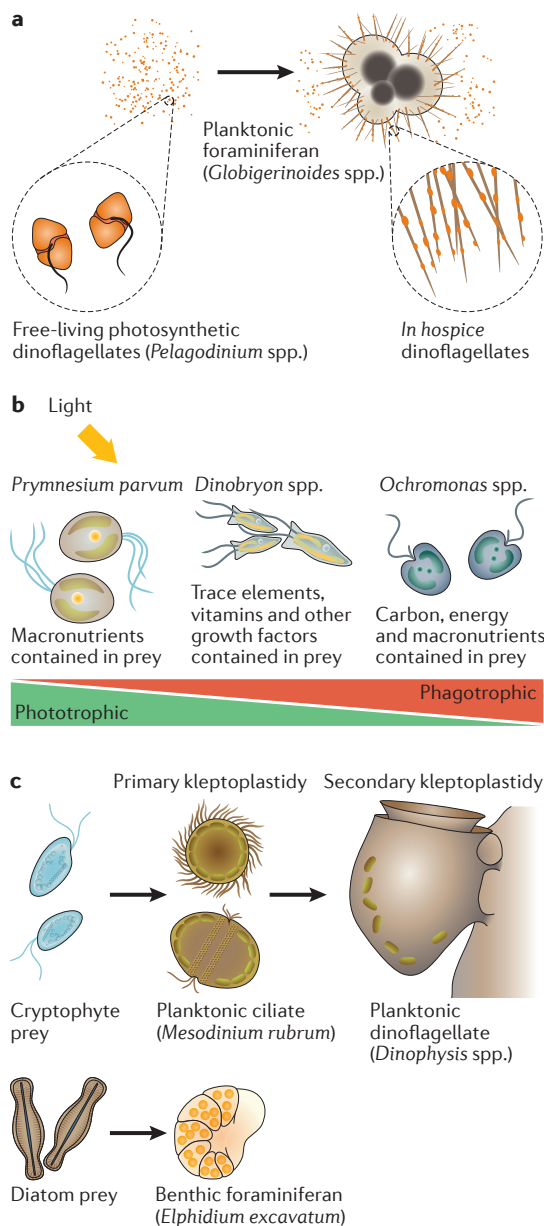
## Microalgal responses to environmental changes.

Microalgae have crucial roles in the fixation and export of carbon from surface waters to the deep ocean, and they form the base of food webs in the most productive regions of the ocean<sup>56</sup>. Global distribution patterns of microalgae can be inferred from satellite-derived assessments of chlorophyll and other pigments<sup>57</sup>, but individual species and taxonomic groups exhibit unique metabolic capabilities that cannot be represented in these larger-scale data. This results in predictive models of primary production and ocean biogeochemistry that lack resolution. Consequently, early transcriptomic and genomic studies have focused strongly on these species with unique capacities and are now bridging this knowledge gap by providing insights into their capabilities and their responses to nutrient availability (which is a major factor determining their activities)<sup>58–60</sup>.

Green algae are one of the few non-parasitic protistan groups with several sequenced genomes, which is a favourable consequence of their small genome sizes and relative ease of culture. Genomic datasets have been instrumental for the alignment and identification of transcriptomic data, which, in turn, have been used to characterize the response of green algae to environmental factors. Approaches are now quite advanced for some freshwater taxa, such as *C. reinhardtii*, which has become a model system for cellular functions and biotechnological applications. For example, transcriptomic, quantitative proteomic and metabolomic data were combined for three strains of *C. reinhardtii* to document their response to nitrogen starvation<sup>61</sup>. There were conserved responses across all strains, which exhibited a differential expression of thousands of genes, with a prioritization for the maintenance of respiration over photosynthesis, and a decreased expression of nitrogen-rich proteins. Studies of synchronous cultures of

## ◀ Figure 2 | The rapidly changing landscape of protistan phylogeny.

DNA sequence information and ultrastructural detail provided by electron microscopy have substantially changed our view of the evolutionary relationships among protistan groups. The system proposed by Whittaker<sup>1</sup>, which prevailed for nearly three decades, grouped eukaryotes into four kingdoms (animals, plants, fungi and protists). The scheme of Woese *et al.*<sup>19</sup> redefined the highest-level organization into domains of life and placed all eukaryotic species within the domain Eukarya. Subsequent revisions (the Sogin system<sup>20</sup>) placed the animals, plants and fungi as only minor evolutionary branches among an enormous diversity of single-celled (protistan) eukaryotic taxa. The basic, redefined structure of the domain emerged approximately a decade ago (Simpson and Roger<sup>21</sup>), with eukaryotes organized into several supergroups. The composition of these supergroups and their evolutionary relationships to one another have begun to stabilize in recent years (Baldauf<sup>22</sup>), but new details within many protistan groups, the placement of many 'orphan' lineages and the relationship of supergroups to each other continue to be proposed and tested (Tekle *et al.*<sup>145</sup> and Adl *et al.*<sup>25</sup>). Our present understanding of protistan evolution (Burki<sup>26</sup>) differs substantially from the scheme proposed by Whittaker. Studies of protistan evolution now incorporate large amounts of genomic and transcriptomic information<sup>27</sup>. Whittaker system adapted from Whittaker, R. H. New concepts of kingdoms of organisms. *Science* **163**, 150–160 (1969). Reprinted with permission from AAAS. Woese scheme reproduced with permission from REF. 146, National Academy of Sciences. Sogin system adapted with permission from REF. 20, Elsevier. Simpson and Roger scheme adapted with permission from REF. 21, Elsevier. Baldauf system adapted with permission from REF. 22, Wiley. Scheme of Tekle *et al.* adapted with permission from REF. 25, Wiley. System of Adl *et al.* adapted from Tekle, Y. I., Wegener Parfrey, L. & Katz, L. A. Molecular data are transforming hypotheses on the origin and diversification of eukaryotes. *Bioscience*, 2009, **59**, 6, 471–481, by permission of Oxford University Press on behalf of the American Institute of Biological Sciences. Burki scheme adapted with permission from REF. 26, Cold Spring Harbor Laboratory Press.



**Figure 3 | Symbiosis and mixotrophy among protists.** Many varieties of mixotrophic nutrition and endosymbiosis exist among protists. **a** | Symbiotic relationships between heterotrophic and photosynthetic protists are extremely common among many large rhizarians in the oceanic plankton<sup>147</sup>. The hosts acquire free-living photosynthetic protists during ontogeny and convert them to intracellular symbionts, as shown here for planktonic foraminiferan hosts and photosynthetic dinoflagellates. Many of these associations involve species-specific relationships between host and endosymbiont. **b** | Prey capture (mostly bacteria, but also other microorganisms) by photosynthetic protists is common among several algal classes and may serve various nutritional needs, including the need for organic carbon for energy and growth (as in some *Ochromonas* spp.), the need to eliminate competing algae (as in *Prymnesium parvum*) or the need to supplement photosynthetic growth with as-yet poorly defined growth factors (as in some *Dinobryon* spp.). These protists span a range of nutritional modes from nearly completely phototrophic (left) to nearly completely heterotrophic (right). These differing strategies for prey ingestion are reflected in the gene expression profiles of these species in response to light and prey<sup>106</sup>. **c** | The retention of functional chloroplasts (kleptoplastidy) from ingested algal prey (shown by arrows in the direction of acquisition) is carried out by numerous heterotrophic dinoflagellates and ciliates, and by a few foraminifera. In one case, cryptophyte chloroplasts retained by the ciliate *Mesodinium rubrum* are also retained by the predators of the ciliate, dinoflagellates within the genus *Dinophysis*. Another example is the retention of diatom chloroplasts by a benthic foraminiferan (for example, *Elphidium excavatum*).

stramenopile *Nannochloropsis oceanica*, in which >7,600 genes (representing >60% of the genome) were shown to cycle in response to the light–dark period<sup>65</sup>.

Diatoms, as major contributors to primary production in the ocean, have become a focus of comparative genomic and transcriptomic analyses to identify genes associated with the metabolic processes that might explain the ecological dominance of these species. Analyses of available diatom genomes, coupled with experiments enabled by transcriptomics data, have highlighted unique aspects of nitrogen metabolism in these species, such as a complete ornithine–urea cycle, which is an important link between nitrogen and inorganic-carbon metabolism<sup>66–68</sup>.

The widely distributed species *Thalassiosira pseudonana* has become a model for linking the genetic and physiological responses of diatoms to environmental cues, in part because of the availability of a sequenced genome<sup>68</sup>. Whole-genome expression profiling of this species has revealed a set of 75 genes that is induced during silicon limitation but not other types of stress<sup>69</sup>. Most of these genes were predicted to have roles in signalling or transport, which might explain the rapid response of diatoms to the changing availability of the element in nature. Furthermore, another analysis used proteomic and transcriptomic data to demonstrate that nitrogen starvation upregulates components of the tricarboxylic acid (TCA) cycle and glycolysis in *T. pseudonana*<sup>70</sup>. The authors of this study speculated that this response might provide carbon and energy for nitrogen assimilation, which would be a significant adaptation in the ocean, in which rapidly

*C. reinhardtii* have also demonstrated sweeping shifts in gene expression (80% of the transcriptome) over a diurnal cycle<sup>62</sup>.

Transcriptomic studies, aided by the availability of sequenced genomes, have also yielded unique insights into the cellular responses of important marine green algae. One such study compared the transcriptomes of the prasinophyte *M. pusilla* grown under phosphorus-replete and phosphorus-deficient conditions<sup>63</sup>, finding that genes that are involved in the acquisition, transport and cellular reallocation of phosphorus are upregulated during phosphorus deficiency. Transcriptomic studies of the prasinophyte *Ostreococcus tauri* revealed extensive rhythmic changes in the gene expression of cultures grown in a light–dark cycle, indicating that the photoperiod exhibits strong transcriptional control<sup>64</sup>. A similar result was observed in the sequenced marine

**Prasinophyte**

A type of green alga (phylum Chlorophyta, supergroup Archaeplastida). Prasinophytes are important primary producers in freshwater and marine ecosystems.

changing nitrogen availability plays a fundamental part in controlling primary productivity. Physiological versatility and responsiveness may also explain the general cellular reorganization that takes place in another diatom, *Phaeodactylum tricoratum*, in response to iron starvation; this reorganization includes the downregulation of iron-rich processes<sup>71</sup>.

Studies of gene expression in diatoms have revealed commonalities that could eventually explain the overall competitive strategy of these protists within natural phytoplankton assemblages, whereas comparative studies among diatom taxa have uncovered many species-specific genes that underlie their unique autecologies. A comparative transcriptomic study of three diatoms (*T. pseudonana*, *Fragilariopsis cylindrus* and *Pseudonitzschia multiseriata*) demonstrated that less than 5% of the >5,500 orthologous genes in these species exhibit the same transcriptional response to nitrate limitation, despite some overall similarities in the cellular responses, such as reduced carbon fixation<sup>72</sup>.

Whole-genome expression profiling has now expanded well beyond several green algae and diatoms, providing valuable datasets for the interpretation of transcriptomic data from other important taxa. Dinoflagellates have been highly desirable targets for these studies because they exhibit an especially wide range of nutritional modes and endosymbiotic relationships, ranging from pure heterotrophy, to obligate endosymbiosis that provides photosynthetic benefits to the heterotrophic host, through to fully integrated and functional chloroplasts<sup>73</sup>. In addition, many species are harmful algal bloom-forming (HAB) species. The very large genome sizes of dinoflagellates have made transcriptome sequencing particularly useful. Recent studies of these species have reported complex transcriptional responses to nutrient limitation or the presence of algicides<sup>74,75</sup>. These complex responses are not unanticipated, given the results obtained with other well-studied algal groups. Moreover, they demonstrate the potential for gaining insight into the cellular processes in ecologically important species with genomes that are too large for modest sequencing projects.

Several non-dinoflagellate HAB species have been investigated to understand their competitive superiority in nature. The pelagophyte *Aureococcus anophagefferens*, another recently sequenced species, has been screened for differential responses of metabolic pathways under three different culture conditions<sup>76</sup>. Genes and pathways that are associated with the uptake and transformation of nitrogen, the hydrolysis of organic phosphorus compounds and the catabolism of organic compounds were upregulated under conditions of low nitrogen, phosphorus or light levels, respectively. The results explain field and laboratory studies indicating that the alga has a considerable capacity for supplementing its nutrition through the uptake of organic compounds<sup>77</sup>. More recently, it was demonstrated that a phosphorus transporter is highly expressed under phosphorus-limiting conditions both in laboratory culture and during a natural bloom of *A. anophagefferens*, thus

exemplifying the power of exploiting gene expression data to interpret the behaviour and nutritional status of natural populations<sup>78</sup>.

Other HAB species use various compounds to slow the growth of competing microalgae (allelopathy)<sup>79</sup>, to deter grazers<sup>80</sup> or to subdue prey<sup>81</sup>. They include such notable compounds as the prymnesins, which are used by the toxic haptophyte *Prymnesium parvum* to subdue and kill prey<sup>82</sup>, and dimethylsulfoniopropionate, which is used by some algae as a deterrent to zooplankton grazing<sup>83</sup>. Transcriptomics and genomics are improving our knowledge both of the sources and biochemical pathways that generate these compounds, and of the environmental conditions that induce their production. For example, powerful neurotoxins called saxitoxins, which affect sodium channel function, were thought to be encoded and produced by bacteria that are associated with dinoflagellates in the genus *Alexandrium*, but recent work has revealed that genes that regulate saxitoxin production are present in the nuclear genomes of several *Alexandrium* spp., which indicates that these toxins might be produced by the protists, either exclusively or in combination with bacterial production<sup>84</sup>. Furthermore, studies of *P. parvum* have hinted at the possible biosynthetic pathways that lead to toxin production in that species<sup>85,86</sup>. Moving forward, the transcriptome assemblies generated for HAB species will provide a targeted reference database for identifying transcriptional patterns, examining genetic variation and contextualizing those signals among populations in nature<sup>87</sup>.

**Metatranscriptomics of natural assemblages of phytoplankton.** Large genome sizes, high species richness, present sequencing depth and cost of sequencing have limited the number of available metatranscriptomic and metagenomic datasets of natural assemblages of eukaryotes. These approaches are now becoming tractable, in part owing to improvements in the reference datasets for gene annotation and in part through focusing on communities with limited species richness<sup>7,88,89</sup>.

Metatranscriptomic studies of natural phytoplankton communities are providing early insights into why diatoms are so successful. A study that investigated the response of diatoms to iron enrichment noted the upregulation of large suites of genes that are associated with nitrate assimilation and synthesis of carbon storage compounds, which indicates that diatoms respond to iron with stimulated nitrogen acquisition and renewed photosynthesis and growth<sup>90</sup>. These findings are consistent with the observation that diatoms are typically the dominant phytoplankton group appearing during blooms in response to iron enrichment experiments in high-nitrate low-chlorophyll regions (HNLC regions)<sup>91</sup>. Similarly, comparative transcriptomics of diatom assemblages from three locations off Antarctica revealed species-specific responses to nutrient status, temperature and light that indicated the unique physiological strategies of the different diatom species that dominated at each of the sites<sup>92</sup>. Furthermore, quantitative transcriptomic studies conducted in coastal Southern Ocean waters have resulted in new insights into the molecular

#### Autecologies

The ecologies of individual species and their interactions with the surrounding environment and co-occurring species.

#### Allelopathy

The production of growth-inhibiting chemicals by one species to target competing species. The term is commonly used in reference to chemical warfare among co-occurring phytoplankton.

#### High-nitrate low-chlorophyll regions

(HNLC regions). Regions of the global ocean in which inorganic nitrogen does not limit phytoplankton growth, and other elements (most notably iron) limit the rate and amount of primary production.



underpinnings of cobalamin and iron limitation and have provided evidence that this limitation is driven by several phytoplankton–bacterium interactions<sup>93</sup>.

Bioinformatic analyses of natural assemblages of phytoplankton have leveraged available microalgal genomes and expressed sequence tag (EST) datasets to map sequence data<sup>89,94</sup>, but gene annotation has been challenging owing to the lack of extensive, well-curated datasets for free-living protists. One recent study showed that approximately 40% of the transcriptome of *T. pseudonana* changed during the photoperiod<sup>95</sup>. However, gene annotation must be improved substantially if we are to gain a better mechanistic understanding of the cellular pathways and processes that are affected by these changes. More than half of the proteins that are predicted in the iron limitation study discussed above<sup>69</sup> had no homology to known proteins.

The application of bioinformatic pipelines using new and more extensive datasets for environmental samples will lead to improved metatranscriptomic analyses. For example, a metatranscriptomic analysis of two dominant diatom species, *Skeletonema costatum* and *Thalassiosira rotula*, showed that both species express core metabolic pathways for nitrogen and phosphorus metabolism, but quantitative metabolic fingerprints (QMFs) of the two species showed opposite transcriptional responses of many genes that are involved in these pathways<sup>96</sup>. Both protists were shown to have homologues that encode nitrate, ammonium and amino acid transporters, but expression of these genes was temporally shifted between the two species. These different capabilities for transporting and transforming phosphorus and nitrogen imply that there is a differential utilization of nutrients (and differential timing of that utilization) between these species, and this example of resource partitioning might explain their co-occurrence.

Using samples that were collected as a part of the Hawaii Ocean Time series (station ALOHA), the expression patterns of three functional groups of phytoplankton (diatoms, dinoflagellates and haptophytes) were investigated; this analysis indicated group-specific differences in metabolic responsiveness to environmental change (FIG. 4), a finding that was supported and refined by the MMETSP dataset<sup>97</sup>. A comparison of the results for two environmental samples yielded a roughly 5-fold increase in sequence read identification using the MMETSP dataset relative to a dataset containing only the available whole-genome reference data for free-living phytoplankton<sup>97</sup> (FIG. 4b). Use of the MMETSP dataset revealed the dominance of dinoflagellate-derived reads in these two samples. However, this dominance was not evident in the earlier analyses, indicating that expansion of the available transcriptomic and genomic information on under-represented protists will improve our understanding of phytoplankton ecology. The QMF of the dinoflagellates was stable over time, whereas haptophytes showed a variable QMF<sup>97</sup>. The cosmopolitan haptophyte *Emiliania huxleyi* was highly represented in the dataset. This species has a set of core genes observed in specimens collected globally (a pangenome)<sup>98</sup>, but also large genetic variability among cultured isolates<sup>99</sup>. It is likely that the genetic variability in this

taxon enables physiological responsiveness to changing environmental conditions, in part explaining the broad geographical distribution of the species.

Sinking-particle flux at station ALOHA during the summer shows a reproducible annual maximum in the export of material from surface waters, which is potentially associated with diatom blooms<sup>100</sup>. Nutrient supply from the mixing of deep, nutrient-rich water into the euphotic zone (for example, through eddy-driven upwelling)<sup>101</sup> is likely to be important for bloom formation. A simulated mixing event (an experiment in which deep water was added to surface water) changed the QMF of the major functional groups of phytoplankton, including expression changes for genes that are involved in carbohydrate, protein and sulfur metabolism, nutrient transport, and photosynthesis<sup>97</sup> (FIG. 4c). Analysis indicated that diatoms have the metabolic plasticity to capitalize on such nutrient pulses, as evidenced by the upregulation of pathways that are associated with photosynthesis and carbon fixation. It is still rare to see the application of transcriptomic data to help explain shifts in primary production and in the taxonomic composition of phytoplankton in response to changing environmental conditions, but this approach constitutes a new and powerful tool in biological oceanography.

**Mixotrophic behaviour and nutrition.** One of the least appreciated but common behaviours of many protists is their ability to combine phototrophic and heterotrophic nutrition. Mixotrophic protists include phototrophic species that also consume prey, and heterotrophic species that retain functional chloroplasts from their prey (kleptoplastidic ciliates, dinoflagellates and some rhizaria).

Heterotrophic nutrition in chloroplast-bearing protists generally involves phagotrophy. Alternatively, or in addition to this, many phototrophic protists are capable of osmotrophy at very high concentrations of labile organic compounds<sup>102</sup>, and osmotrophy may provide planktonic algae with specific organic substances required in trace amounts (for example, vitamins)<sup>103</sup>. However, there is no convincing evidence that osmotrophy alone can support measurable population growth of algae in marine pelagic ecosystems in which concentrations of organic compounds are exceedingly low.

Phagotrophy is widespread among the major taxonomic groups of phytoplankton, including the chrysophytes, cryptophytes, dinoflagellates, haptophytes and green algae. These mixotrophic groups can use predation to satisfy various nutritional requirements (the need for carbon, energy, major nutrients, trace metals or vitamins), and they have very variable phototrophic and heterotrophic capabilities (FIG. 3b). Experimental studies have attempted to resolve the ecological strategies and resulting competitive advantages of mixotrophy in these species and have tried to incorporate those properties into food web models<sup>104,105</sup>, but we currently have a poor understanding of this behaviour.

Comparative transcriptomic studies have begun to shed light on the physiological abilities, and thereby ecological strategies, of mixotrophic algae.

#### Quantitative metabolic fingerprints

(QMFs). Comparisons of the relative expression levels of genes or gene families for a species under different environmental conditions or for different species under the same conditions. The fingerprints can indicate specific pathways that are upregulated or downregulated, and can thus provide insight into the metabolic responses of the organism.

#### Kleptoplastic

Able to ingest and partially digest photosynthetic prey, but retain the chloroplasts from the prey in a functional state, usually for periods of days to a few weeks.

#### Ciliates

Members of the phylum Ciliophora (supergroup Alveolata), a monophyletic group of heterotrophic and mixotrophic (kleptoplastidic) protists characterized by the presence of cilia that are used for motility and feeding. The ciliates are major consumers of phytoplankton, bacteria and other microorganisms.

#### Phagotrophy

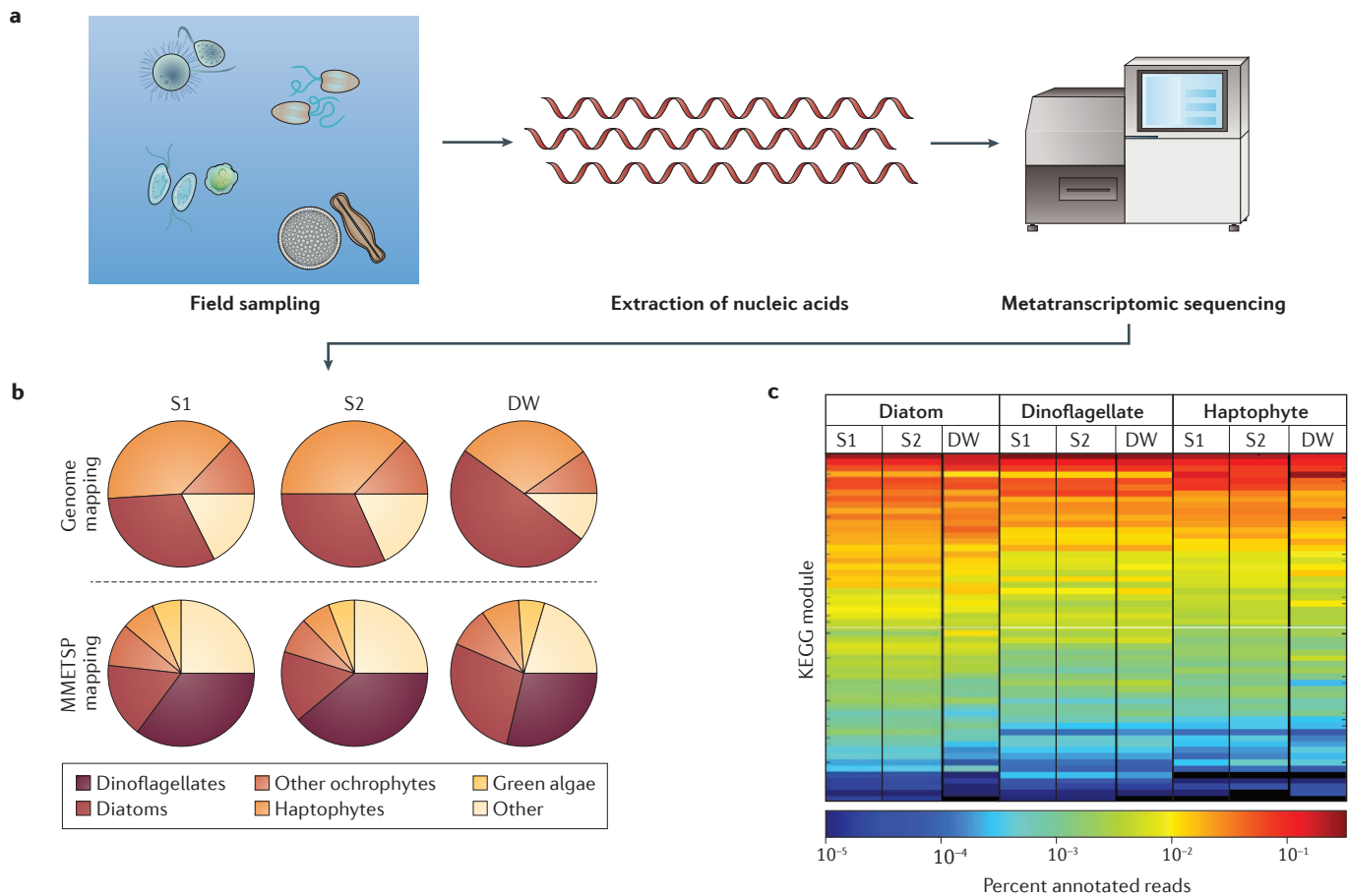
A mode of nutrition that is characterized by the engulfment and digestion of particulate material, usually microbial prey.

#### Osmotrophy

A mode of nutrition that is characterized by the absorption and utilization of dissolved organic compounds.

#### Chrysophytes

Small, flagellated algal protists of the class Chrysophyceae (supergroup Stramenopiles), which contains photosynthetic, heterotrophic and mixotrophic species. Chloroplast-bearing species were previously referred to as golden-brown algae, an imprecise term no longer used by specialists.



**Figure 4 | Improved gene identification and understanding of phytoplankton functional groups. a, b** | Metatranscriptomes from a natural phytoplankton assemblage (part **a**) were mapped to available genomes of free-living algae (part **b**; top row) and to a custom-built Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP)-based transcriptome dataset (part **b**; bottom row); mapping was carried out for two near-surface *in situ* assemblage samples, S1 and S2, and also for a sample that was cultured with deep water added (DW). In comparison to the use of the available whole genomes, the MMETSP dataset enabled a

substantial increase in the ability to assign genes to specific phytoplankton groups<sup>97</sup>. **c** | A heat map was generated of the quantitative metabolic fingerprints (QMFs) of the major eukaryotic phytoplankton functional groups (diatoms, dinoflagellates and haptophytes) in the nanoplankton to microplankton size class (>5 μm). Gene identity was based on mapping to MMETSP KEGG orthology and revealed group-specific responses to the addition of nutrient-rich deep water to surface-water communities. Parts **b** and **c** adapted with permission from REF. 97, National Academy of Sciences.

A comparative study of two species of chrysophyte (*Dinobryon* sp. and *Ochromonas* sp.) and a haptophyte (*P. parvum*) showed that these species possess different phototrophic and heterotrophic abilities (FIG. 3b) and exhibit corresponding differences in gene expression<sup>106</sup>. *Ochromonas* sp., a predominantly heterotrophic species<sup>107</sup>, grew rapidly in the presence of bacteria regardless of light conditions, whereas its growth in the absence of prey was a fraction of its heterotrophic growth potential. Expression levels of the rRNA genes of the alga were relatively unchanged between light and dark, as long as bacterial prey was present, which indicates the preference of this species for heterotrophy. By contrast, *P. parvum* readily immobilized and killed ciliate prey, but the expression of rRNA genes was strongly upregulated in the light compared with levels in the dark, consistent with a physiology that is more dependent on phototrophy than on heterotrophy. The upregulation of rRNA genes by *Dinobryon* sp. in the light was similar to the response of *P. parvum*,

revealing a strong dependence on phototrophy even though this species also readily consumes prey. The comparative transcriptomic approach is an important step towards understanding the ecology of these mixotrophic algae in nature.

On the other end of the trophic spectrum, many heterotrophic dinoflagellates and ciliates, as well as a few foraminifera, supplement heterotrophic nutrition through organic-carbon and energy acquisition by the retention of functional chloroplasts from ingested photosynthetic prey<sup>108,109</sup>. Genetic studies of these kleptoplastidic mixotrophs (FIG. 3c) are benefiting from the growth of public transcriptomic reference databases to aid gene annotation. It is now clear that the ciliate *Mesodinium rubrum* (*Myrionecta rubra*) can maintain transcriptionally active chloroplasts that are acquired from its microalgal prey<sup>110</sup>. Moreover, a comparison of gene expression in the photosynthetic prey in the free-living state and in the host-sequestered state indicate that specific metabolic pathways are maintained in the sequestered state

(notably, photosynthesis and oxidative-stress reduction)<sup>111</sup>. Future efforts to decipher the molecular communication between host and prey organelles will rely heavily on investigations of gene expression.

**Heterotrophic protists: a neglected majority.** Predator–prey and parasite–host relationships that involve protists are crucial for the food webs of aquatic and soil ecosystems (FIG. 1). The Tara Oceans expedition collected thousands of plankton samples for analysis and thus generated a vast amount of data concerning marine protists, network analysis of which has demonstrated the tremendous diversity and trophic connectivity of these organisms<sup>112</sup>. Findings that reveal the high diversity and connectivity of eukaryotic parasitic taxa, in particular, are changing our view of the complexity of food web structures within these communities. Nevertheless, there have been very few genetic studies of parasitic or free-living heterotrophic protists aimed at understanding these food web interactions, owing mostly to the lack of well-curated culture collections for heterotrophs. Consequently, knowledge of the metabolic processes that are involved in heterotrophic nutrition has been gleaned mostly from studies of non-marine model systems or from studies with other motivations, such as understanding evolution or pathogenicity.

A few transcriptomic studies focused on identifying the genes involved in phagocytosis and examining how metabolism is affected by this process, and these studies have begun to provide some insight into the response of protists to prey (FIG. 5a). Nearly 75% of genes in slime moulds of the genus *Dictyostelium* were found to be differentially expressed in the presence or absence of prey<sup>113</sup>, which indicates that these species have a complex response to phagotrophic nutrition, including changes in the expression of genes that are responsible for prey recognition at the membrane, adhesion, transcription regulation and sterol metabolism (the latter seems to contribute to phagosomal membrane formation). In the phagotrophic haptophyte *P. parvum*, predation by the alga resulted in upregulation of genes that are involved in fatty acid oxidation and the TCA cycle<sup>114</sup>, whereas genes that are related to inorganic-nitrogen uptake and transformation and iron uptake were substantially downregulated when prey was present, implying that nitrogen and iron are obtained from ingested prey.

Genomic and transcriptomic investigations of protistan human parasites have uncovered factors that are involved in host invasion, cytotoxicity, metabolism, virulence and immunosuppression or immunoevasion<sup>115</sup>. By comparison, little is known about gene expression in marine parasitoids, but one recent study of an *Amoebophrya* sp. (the genus is a major parasitoid group, members of which infect and kill numerous phytoplankton) has indicated the probable role of lectins in attachment to the host<sup>116</sup>. Interestingly, in another study<sup>46</sup> lectins seemed to be involved in the establishment of endosymbiotic associations between large rhizaria and photosynthetic dinoflagellates.

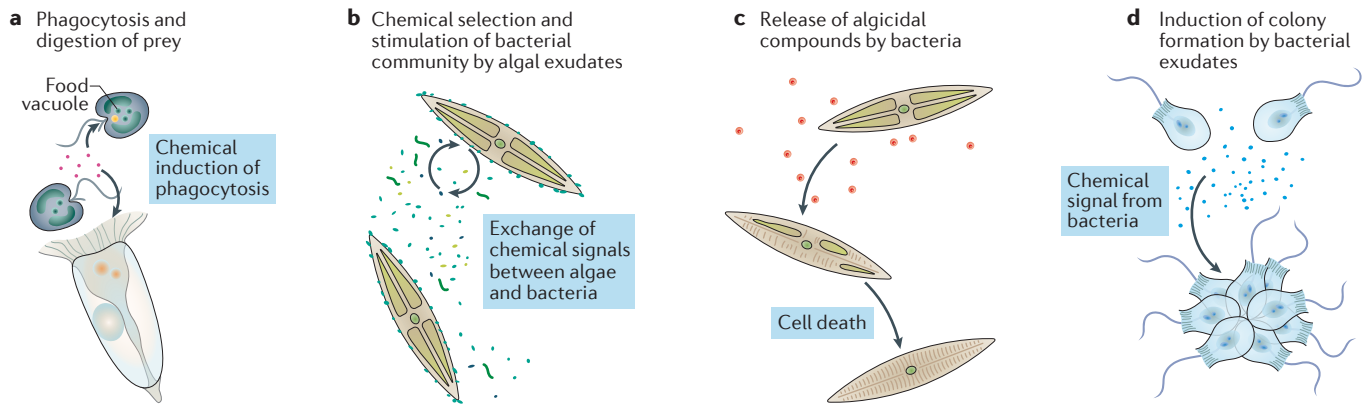
Beyond these meagre insights concerning marine heterotrophic protists, there has been little examination of gene expression in heterotrophic species. The heterotrophic dinoflagellate *Oxyrrhis marina* has recently emerged as a model system for studying the genetics of free-living heterotrophic protists, but transcriptomic analyses of this species have focused not on processes such as phagotrophy but rather on the evolutionary origin of dinoflagellate features<sup>117</sup> and on rhodopsin expression in response to starvation<sup>118</sup>.

**Protist–bacterium interactions.** Whereas work on predator–prey, host–symbiont and host–parasitoid relationships is only just beginning, transcriptomic investigations of phototrophic protists and their associated bacterial assemblages have rapidly advanced our understanding of protist–bacterium interactions<sup>119</sup> (FIG. 5b–d). The phycosphere — the chemical and physical micro-environment that surrounds phytoplankton cells — has been a topic of research for many years<sup>120</sup>, but studies of the diversity and gene expression of protistan hosts and their co-existing bacteria are now revealing the true nature of these associations (FIG. 5b). Algae impose strong selection on the bacterial community<sup>121,122</sup>, nurturing the growth of specific bacterial taxa that in turn provide growth substances required by the protist, resulting in mutualistic metabolic relationships<sup>88,123</sup>. The ecological roles of these interactions for the protist range from nitrogen acquisition, such as in associations between nitrogen-fixing cyanobacteria and algae<sup>124</sup>, to vitamin acquisition in many microalgae<sup>125</sup>.

The ability of bacteria to directly affect gene expression in some diatoms and dinoflagellates has also been demonstrated<sup>126,127</sup>. For example, one investigation showed that in the toxic dinoflagellate *Alexandrium tamarense*, S-adenosylmethionine synthetase (SAMS)-encoding genes are downregulated in the presence of bacteria<sup>127</sup>. The authors noted the similarity between the response of the dinoflagellate to bacteria and the relationship reported between *Amoeba proteus* and its endosymbiotic bacterium, in which the presence of the endosymbiont downregulates host genes that encode SAMS<sup>128</sup>. This activity seems to make the host dependent on SAMS produced by the bacterium, and this possibly solidifies the symbiotic relationship by linking the survival and growth of the host to its bacterial associate.

Another study noted the production of algicidal compounds by a bacterial associate of the haptophyte *E. huxleyi* and, furthermore, determined that this was elicited by compounds that are produced by senescent algal cells<sup>129</sup> (FIG. 5c). The investigators proposed a model whereby the production of algicidal compounds converted a normally mutualistic relationship into a parasitic one. There is a high probability that many such complex and specific interactions between bacteria and marine protists await discovery.

Colony formation in the choanoflagellate *Salpingoeca rosetta* has been shown to be stimulated (or inhibited) by specific lipids produced by co-occurring bacteria<sup>130</sup> (FIG. 5d), a finding that may have implications for the development of multicellularity in the supergroup



**Figure 5 | Interactions between protists and other microorganisms.** A myriad of interactions between protistan species and other organisms await discovery and characterization. Transcriptomic studies are now beginning to reveal changes in gene expression and the metabolic pathways that are associated with some of these interactions. **a** | Phagotrophy of a prey organism is induced by chemicals that are contained in the prey<sup>114</sup> and can lead to changes in gene expression, such as upregulation of receptor genes and downregulation of genes involved in the uptake of nutrients that can be obtained from the prey (instead of by direct uptake from the environment). **b** | Chemical communication between bacteria and algae can give rise to species-specific mutualistic associations<sup>121</sup>. **c** | Chemical warfare between algae and some co-occurring bacteria can result in algal death<sup>129</sup>. **d** | Chemical signalling can alter the life cycle of some protists. In one example, the formation of a colony by protists is induced (or inhibited) in response to lipids that are produced by co-existing bacteria<sup>130</sup>.

Opisthokonta, which also contains the animals and fungi. Interestingly, a colony-inducing sulfonolipid produced by the bacterium *Algoriphagus machipongonensis* resembles molecules that are produced by plants, animals and fungi and that have roles in signal transmission.

Many interactions undoubtedly also exist between marine protists and larger, multicellular organisms, and the molecular interactions that control competition, predation, parasitism, mutualism and commensalism are just beginning to be characterized. Likewise, the interplay between protists and viruses is particularly poorly understood. Although viruses of eukaryotic algae have been known for many years<sup>131</sup>, and more recently viruses of heterotrophic protists have also been characterized<sup>132</sup>, the factors that control protist susceptibility and resistance, and the overall ecological role of these interactions, are poorly understood.

**Outlook and conclusions**

Studies of protistan biology using traditional approaches have broadly framed the contributions of these species to important biogeochemical processes. However, physiological details about species and functional groups of many eukaryotes have thus far proved insufficient to produce global models with strong predictive power. The next great challenge is the integration of new information regarding various aspects of protistan biology, from the genetic potential of these organisms to the behaviours of species, populations and ecosystems as a whole. In-depth genetic investigations, enabled by large-scale genomic and transcriptomic datasets, are beginning to meet that challenge. These approaches provide a fundamental mechanistic understanding of the metabolic potential of protists and their responses to specific conditions and interactions. Transcriptomes

are also providing a much-needed reference dataset for the interpretation of genetic data obtained from the massive field studies that have recently been completed<sup>4,18</sup>. What is more, the promise of this approach is stimulating genomic and transcriptomic projects on species from lineages not yet represented among sequenced (or even cultured) eukaryotes.

Recent efforts to obtain extensive genetic information about free-living marine protists have generated a much-expanded database of protistan transcriptomes, facilitating genetic studies on a broad diversity of these species. In addition, the transcriptomes of a few freshwater protists (for example, glaucophytes) were completed among the 396 strains that were sequenced by the MMETSP project, because many protistan lineages contain species from marine, freshwater and terrestrial environments. Therefore, the discoveries that are made possible by new datasets will have a substantial impact on investigations into a broad spectrum of free-living protists. These transcriptomes also provide a starting point for comparative studies that aim to understand the genetic foundations of adaptation to marine, freshwater and terrestrial existence.

**The need for functional genomics.** Eukaryotic transcriptomics (and, to some degree, metatranscriptomics) have expanded rapidly in recent years. RNA depletion and poly-A selection (the latter approach misses non-coding RNA and non-poly-A RNA transcripts) have been used effectively to lower sequencing costs and improve sequencing depth. However, transcriptomes are not genomes. Transcriptomes recover only the genes that are expressed at the time of sampling and so do not define all the genetic potential of a species. This issue can be addressed to some extent by combining multiple transcriptomes from a strain that is cultured under



different environmental conditions, yielding assemblies that provide a more complete picture of the genetic potential of the taxon<sup>86</sup>. Even so, gene expression does not necessarily imply protein function, owing to the potential for post-transcriptional and post-translational processing.

In addition, most protistan species have not yet been cultured and are therefore unavailable for transcriptome sequencing. Metatranscriptomes are a partial solution to this problem, but as well as sequencing costs, gene annotation remains a problem (for example, up to 75% of genes cannot yet be ascribed a function<sup>47,96</sup>), and many genes cannot even be assigned to a particular taxon. Sequenced genomes, and the reference databases that they provide, will therefore be exceptionally important in the future. They are fundamental for the alignment and annotation of datasets derived from transcriptomes and metatranscriptomes, and they enable analyses using in-depth functional genomics methods.

Genomic studies are becoming increasingly feasible with present-day sequencing power, even for species with large genomes, such as dinoflagellates<sup>42</sup>. Consequently, transcriptomic analyses for species with sequenced genomes are appearing for species other than green algae and diatoms<sup>133,134</sup>. Beyond individual genomes, eukaryotic metagenomics and the reconstruction of protistan genomes from these datasets are currently feasible for only a limited number of species with small genomes, for samples from environments with very low species richness (for example, some extreme environments) or in only exceptionally well-funded projects. Even when these approaches are feasible, the interpretation of their results will be limited by the lack of suitable reference databases for assigning taxonomy and for gene annotation. These studies will continue to lag behind similar studies of prokaryotes.

**Future directions.** Future applications of transcriptomics to the study of protistan biology will most probably integrate other technologies that are now being used for prokaryotes. Most notably, single-cell sorting techniques and similar methods will be essential for isolating, analysing and culturing protists from environments in which

little is known about the distribution of many unique, uncultured taxa<sup>135</sup>. This technology can also be used to characterize the viral, bacterial and archaeal assemblages that are associated with particular protistan strains and species<sup>136,137</sup>, facilitating experimental investigation of the nature of these ecological interactions.

Although transcriptomics provides a robust starting point for gaining a better understanding of free-living marine protists, there are limitations to the approach, as noted above. Moreover, genomic features that regulate transcription, such as promoters and methylation, are not well understood<sup>138</sup>, nor are post-translational modifications. In addition, the total amount of RNA per cell is often assumed to be fairly constant, but this is not necessarily true. The existence of different levels of total RNA in any two cells to be compared (transcriptional amplification) can lead to erroneous conclusions regarding relative gene expression levels<sup>139</sup>. Some normalization strategies being developed include, but are not limited to, the use of exogenous standards and cell counts, or normalization to stable reference genes. For these and other reasons, it will be necessary to combine transcriptomic, proteomic, metabolomic and gene manipulation studies of marine protists to validate any inferences made from studies of gene expression<sup>140,141</sup>.

Collectively, the insights gained from this research on the evolution, biochemical capabilities and activities of ecologically important protistan species will be widely felt. For example, genomic sequencing of the tsetse fly and its associated protist, which causes trypanosomiasis<sup>142</sup>, enhanced our knowledge of the interplay between these two species and presented possible research avenues for the prevention of human disease. Similarly, the study of free-living protists will advance our understanding of how environmental change might lead to shifts in food web structure and ecosystem function. In turn, this knowledge will guide the design of biogeochemical models to better predict these outcomes and to help determine ecosystem-level management efforts on local, regional and global scales<sup>143,144</sup>. In recognition of these possibilities, it is true to say that molecular ecological studies of marine protists have finally come of age.

- Whittaker, R. H. New concepts of kingdoms of organisms. *Science* **163**, 150–160 (1969).
- Ohtsuka, S., Suzuki, T., Horiguchi, T., Suzuki, N. & Not, F. (eds) *Marine Protists* (Springer, 2015).
- Worden, A. Z. *et al.* Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347**, 1257594 (2015).  
**An overview of protistan contributions to biogeochemical cycles.**
- de Vargas, C. *et al.* Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605 (2015).  
**Data obtained from the Tara Oceans expedition expands our knowledge of the global diversity of protists.**
- Keeling, P. J. *et al.* 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 (2014).  
**An article delineating the contribution of the MMETSP to expanding the global databases for gene discovery and annotation in protists.**
- del Campo, J. *et al.* The others: our biased perspective of eukaryotic genomes. *Trends Ecol. Evol.* **29**, 252–259 (2014).
- Cuvelier, M. L. *et al.* Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc. Natl Acad. Sci. USA* **107**, 14679–14684 (2010).
- Caron, D. A. Towards a molecular taxonomy for protists: benefits, risks and applications in plankton ecology. *J. Eukaryot. Microbiol.* **60**, 407–413 (2013).
- Fenchel, T. & Finlay, B. J. The ubiquity of small species: patterns of local and global diversity. *Bioscience* **54**, 777–784 (2004).
- Foissner, W. Biogeography and dispersal of micro-organisms: a review emphasizing protists. *Acta Protozool.* **45**, 111–136 (2006).
- Guillou, L. *et al.* Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environ. Microbiol.* **10**, 3349–3365 (2008).
- Shalchian-Tabrizi, K., Kausserud, H., Massana, R., Klaveness, D. & Jakobsen, K. S. Analysis of environmental 18S ribosomal RNA sequences reveals unknown diversity of the cosmopolitan phylum Telonemia. *Protist* **158**, 173–180 (2007).
- Caron, D. A. & Countway, P. D. Hypotheses on the role of the protistan rare biosphere in a changing world. *Aquat. Microb. Ecol.* **57**, 227–238 (2009).
- Logares, R. *et al.* Patterns of rare and abundant marine microbial eukaryotes. *Curr. Biol.* **24**, 813–821 (2014).
- Lynch, M. D. J. & Neufeld, J. D. Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* **13**, 217–229 (2015).
- Caron, D. A., Worden, A. Z., Countway, P. D., Demir, E. & Heidelberg, K. B. Protists are microbes too: a perspective. *ISME J.* **3**, 4–12 (2009).
- Cohen, L., Alexander, H. & Brown, C. T. Marine Microbial Eukaryotic Transcriptome Sequencing Project, re-assemblies. *Figshare* <https://dx.doi.org/10.6084/m9.figshare.3840153.v1> (2016).
- Pernice, M. C. *et al.* Global abundance of planktonic heterotrophic protists in the deep ocean. *ISME J.* **9**, 782–792 (2015).
- Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria and Eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
- Sogin, M. L. Early evolution and the origin of eukaryotes. *Curr. Opin. Genet. Dev.* **1**, 457–463 (1991).
- Simpson, A. G. B. & Roger, A. J. The real 'kingdoms' of eukaryotes. *Curr. Biol.* **14**, R693–R696 (2004).

22. Baldauf, S. L. An overview of the phylogeny and diversity of eukaryotes. *J. Systemat. Evol.* **46**, 263–273 (2008).
23. Pawlowski, J. & Burki, F. Untangling the phylogeny of amoeboid protists. *J. Eukaryot. Microbiol.* **56**, 16–25 (2009).
24. Burki, F. *et al.* Phylogenomics reshuffles the eukaryotic supergroups. *PLoS ONE* **8**, E790 (2007).
25. Adl, S. M. *et al.* The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59**, 429–514 (2012).
26. Burki, F. The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* **6**, a016147 (2014).
27. Grant, J. R. & Katz, L. A. Building a phylogenomic pipeline for the eukaryotic tree of life — addressing deep phylogenies with genome-scale data. *PLoS Curr.* <http://dx.doi.org/10.1371/currents.tol.c24b6054aebf3602748ac042ccc8f2e9> (2014).
28. Koonin, E. V. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol.* **11**, 209 (2010).
29. Barberà, M. J., Ruiz-Trillo, I., Leigh, J., Hug, L. A. & Roger, A. J. In *Origin of Mitochondria and Hydrogenosomes* (eds Martin, W. F. & Müller, M.) 239–275 (Springer, 2007).
30. Gould, S. B., Waller, R. F. & McFadden, G. I. Plastid evolution. *Annu. Rev. Plant Biol.* **59**, 491–517 (2008).
31. Keeling, P. J. The number, speed, and impact of plastid endosymbioses on eukaryotic evolution. *Annu. Rev. Plant Biol.* **64**, 583–607 (2013).
32. Deschamps, P. & Moreira, D. Reevaluating the green contribution to diatom genomes. *Genome Biol. Evol.* **4**, 683–688 (2012).
33. Burki, F. *et al.* Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. *Eukaryot. Cell* **13**, 246–255 (2014).
34. Slamovits, C. H. & Keeling, P. J. Plastid-derived genes in the non-photosynthetic alveolate *Oxymyces marina*. *Mol. Biol. Evol.* **25**, 1297–1306 (2008).
35. Moore, C. E. & Archibald, J. M. Nucleomorph genomes. *Annu. Rev. Genet.* **43**, 251–264 (2009).
36. Curtis, B. A. *et al.* Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492**, 59–65 (2012).
37. Nakayama, T. & Ishida, K.-I. Another acquisition of a primary photosynthetic organelle in seaweed *Paulinella chromatophora*. *Curr. Biol.* **19**, R284–R285 (2009).
38. Nowack, E. C. *et al.* Endosymbiotic gene transfer and transcriptional regulation of transferred genes in *Paulinella chromatophora*. *Mol. Biol. Evol.* **28**, 407–422 (2011).
39. Nowack, E. C. & Grossman, A. R. Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. *Proc. Natl Acad. Sci. USA* **109**, 5340–5345 (2012).
40. Nakayama, T. & Archibald, J. M. Evolving a photosynthetic organelle. *BMC Biol.* **10**, 34 (2012).
41. Estrela, S., Kerr, B. & Morris, J. J. Transitions in individuality through symbiosis. *Curr. Opin. Microbiol.* **31**, 191–198 (2016).
42. Lin, S. *et al.* The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* **350**, 691–694 (2015).
43. Bayer, T. *et al.* *Symbiodinium* transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. *PLoS ONE* **7**, e35269 (2012).
44. Biard, T. *et al.* *In situ* imaging reveals the biomass of large protists in the global ocean. *Nature* **532**, 504–507 (2016).
45. Decelle, J. *et al.* An original mode of symbiosis in open ocean plankton. *Proc. Natl Acad. Sci. USA* **109**, 18000–18005 (2012).
46. Balzano, S. *et al.* Transcriptome analyses to investigate symbiotic relationships between marine protists. *Front. Microbiol.* **6**, 98 (2015).
47. Pillet, L. & Pawlowski, J. Transcriptome analysis of foraminiferan *Elphidium margaritaceum* questions the role of gene transfer in kleptoplastidy. *Mol. Biol. Evol.* **30**, 66–69 (2013).
48. Chen, M. M., Chory, J. & Fankhauser, C. Light signal transduction in higher plants. *Annu. Rev. Genet.* **38**, 87–117 (2004).
49. Falciatore, A. & Bowler, C. The evolution and function of blue and red light photoreceptors. *Curr. Top. Dev. Biol.* **68**, 317–350 (2005).
50. Rodríguez-Romero, J., Hedtke, M., Kastner, C., Müller, S. & Fischer, R. Fungi, hidden in soil or up in the air: light makes a difference. *Annu. Rev. Microbiol.* **64**, 585–610 (2010).
51. Fortunato, A. E. *et al.* Diatom phytochromes reveal the existence of far-red light-based sensing in the ocean. *Plant Cell* **28**, 616–628 (2016).
52. Worden, A. Z. *et al.* Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* **324**, 268–272 (2009).
53. Duanmu, D. *et al.* Marine algae and land plants share conserved phytochrome signaling systems. *Proc. Natl Acad. Sci. USA* **111**, 15827–15832 (2014).
54. Moreau, H. *et al.* Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biol.* **13**, 1–16 (2012).
55. Allen, A. E. *et al.* Evolution and functional diversification of fructose biphosphate aldolase genes in photosynthetic marine diatoms. *Mol. Biol. Evol.* **29**, 367–379 (2012).
56. Armbrust, E. The life of diatoms in the world's oceans. *Nature* **459**, 185–192 (2009).
57. Farikou, O. *et al.* Inferring the seasonal evolution of phytoplankton groups in the Senegalo-Mauritanian upwelling region from satellite ocean-color spectral measurements. *J. Geophys. Res.: Oceans* **120**, 6581–6601 (2015).
58. Dyhrman, S. T. *et al.* The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLoS ONE* **7**, e33768 (2012).
59. Gobler, C. J. *et al.* Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc. Natl Acad. Sci. USA* **108**, 4352–4357 (2011).
60. Shrestha, R. P. *et al.* Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. *BMC Genomics* **13**, 499 (2012).
61. Schmollinger, S. *et al.* Nitrogen-sparing mechanisms in *Chlamydomonas* affect the transcriptome, the proteome, and photosynthetic metabolism. *Plant Cell* **26**, 1410–1435 (2014).
62. Zones, J. M., Blaby, I. K., Merchant, S. S. & Umen, J. G. High-resolution profiling of a synchronized diurnal transcriptome from *Chlamydomonas reinhardtii* reveals continuous cell and metabolic differentiation. *Plant Cell* **27**, 2743–2769 (2015).
63. Whitney, L. P. & Lomas, M. W. Growth on ATP elicits a P-stress response in the picoeukaryote *Micromonas pusilla*. *PLoS ONE* **11**, e0155158 (2016).
64. Monnier, A. *et al.* Orchestrated transcription of biological processes in the marine picoeukaryote *Ostreococcus* exposed to light/dark cycles. *BMC Genomics* **11**, 192 (2010).
65. Poliner, E. *et al.* Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1179 under light/dark cycles. *Plant J.* **85**, 1097–1113 (2015).
66. Allen, A. E., Vardi, A. & Bowler, C. An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Curr. Opin. Plant Biol.* **9**, 264–273 (2006).
67. Allen, A. E. *et al.* Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **473**, 205–207 (2011).
68. Armbrust, E. V. *et al.* The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004).
69. Mock, T. *et al.* Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. *Proc. Natl Acad. Sci. USA* **105**, 1579–1584 (2008).
70. Hockin, N. L., Mock, T., Mulholland, F., Kopriva, S. & Malin, G. The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants. *Plant Physiol.* **158**, 299–312 (2012).
71. Allen, A. E. *et al.* Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl Acad. Sci. USA* **105**, 10438–10443 (2008).
72. Bender, S. J., Durkin, C. A., Berthiaume, C. T., Morales, R. L. & Armbrust, E. V. Transcriptional responses of three model diatoms to nitrate limitation of growth. *Front. Mar. Sci.* **1**, 3 (2014).
73. Hackett, J. D., Anderson, D. M., Erdner, D. L. & Bhattacharya, D. Dinoflagellates: a remarkable evolutionary experiment. *Am. J. Bot.* **91**, 1523–1534 (2004).
74. Cooper, J. T., Sinclair, G. & Wawrik, B. Transcriptome analysis of *Scrippsiella trochoidea* CCMP 3099 reveals physiological changes related to nitrate depletion. *Front. Microbiol.* **7**, 659 (2016).
75. Guo, R., Wang, H., Suh, Y. S. & Ki, J.-S. Transcriptomic profiles reveal the genome-wide responses of the harmful dinoflagellate *Cochlodinium polykrikoides* when exposed to the algicidal copper sulfate. *BMC Genomics* **17**, 29 (2016).
76. Frischkorn, K. R., Harke, M. J., Gobler, C. J. & Dyhrman, S. T. *De novo* assembly of *Aureococcus anophagefferens* transcriptomes reveals diverse responses to the low nutrient and low light conditions present during blooms. *Front. Microbiol.* **5**, 375 (2014).
77. Gobler, C. J., Boneillo, G. E., Debenham, C. J. & Caron, D. A. Nutrient limitation, organic matter cycling, and plankton dynamics during an *Aureococcus anophagefferens* bloom. *Aquat. Microb. Ecol.* **35**, 31–43 (2004).
78. Wurch, L. L., Gobler, C. J. & Dyhrman, S. T. Expression of a xanthine permease and phosphate transporter in cultures and field populations of the harmful alga *Aureococcus anophagefferens*: tracking nutritional deficiency during brown tides. *Environ. Microbiol.* **8**, 2444–2457 (2014).
79. Qiu, X. *et al.* Allelopathy of the raphidophyte *Heterosigma akashiwo* against the dinoflagellate *Akashiwo sanguinea* is mediated via allelochemicals and cell contact. *Mar. Ecol. Prog. Ser.* **446**, 107–118 (2012).
80. Smetacek, V. A watery arms race. *Nature* **411**, 745 (2001).
81. Rimmel, E. J. & Hambright, K. D. Toxin-assisted micropredation: experimental evidence shows that contact micropredation rather than exotoxicity is the role of *Prymnesium* toxins. *Ecol. Lett.* **15**, 126–132 (2012).
82. Manning, S. R. & La Claire, J. W. *Prymnesium*: toxic metabolites of the golden alga, *Prymnesium parvum* Carter (Haptophyta). *Mar. Drugs* **8**, 678–704 (2010).
83. Strom, S., Wolfe, G. V., Slajer, A., Lambert, S. & Clough, J. Chemical defenses in the microplankton II: inhibition of protist feeding by  $\beta$ -dimethylsulfoniopropionate (DMSP). *Limnol. Oceanogr.* **48**, 230–237 (2003).
84. Stüken, A. *et al.* Discovery of nuclear-encoded genes for the neurotoxin saxitoxin in dinoflagellates. *PLoS ONE* **6**, e20096 (2011).
85. Beszteri, S. *et al.* Transcriptomic response of the toxic prymnesiophyte *Prymnesium parvum* (N. Carter) to phosphorus and nitrogen starvation. *Harmful Algae* **18**, 1–15 (2012).
86. Liu, Z. *et al.* Changes in gene expression of *Prymnesium parvum* due to nitrogen and phosphorus limitation. *Front. Microbiol.* **6**, 631 (2015).
87. McLean, T. I. "Eco-omics": a review of the application of genomics, transcriptomics, and proteomics for the study of the ecology of harmful algae. *Microb. Ecol.* **65**, 901–915 (2013).
88. Delmont, T. O., Eren, A. M., Vineis, J. H. & Post, A. F. Genome reconstructions indicate the partitioning of ecological functions inside a phytoplankton bloom in the Amundsen Sea, Antarctica. *Front. Microbiol.* **6**, 1090 (2015).
89. Ottesen, E. A. *et al.* Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proc. Natl Acad. Sci. USA* **110**, E488–E497 (2013).
90. Marchetti, A. *et al.* Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl Acad. Sci. USA* **109**, E317–E325 (2012).
91. Boyd, P. W. *et al.* A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* **407**, 695–702 (2000).
92. Pearson, G. A. *et al.* Metatranscriptomes reveal functional variation in diatom communities from the Antarctic Peninsula. *ISME J.* **9**, 2275–2289 (2015). **A metatranscriptomic study of phytoplankton communities off the Antarctic peninsula.**
93. Bertrand, E. M. *et al.* Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proc. Natl Acad. Sci. USA* **112**, 9938–9943 (2015).
94. Aylward, F. O. *et al.* Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proc. Natl Acad. Sci. USA* **112**, 5443–5448 (2015).
95. Ashworth, J. *et al.* Genome-wide diel growth state transitions in the diatom *Thalassiosira pseudonana*. *Proc. Natl Acad. Sci. USA* **110**, 7518–7523 (2013).

96. Alexander, H., Jenkins, B. D., Rynearson, T. A. & Dyhrman, S. T. Metatranscriptome analyses indicate resource partitioning between diatoms in the field. *Proc. Natl Acad. Sci. USA* **112**, E2182–E2190 (2015). **The application of transcriptomic approaches help to identify examples of resource competition and to understand its effects within natural phytoplankton assemblages.**
97. Alexander, H. *et al.* Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proc. Natl Acad. Sci. USA* **112**, E5972–E5979 (2015).
98. von Dassow, P. *et al.* Life-cycle modification in open oceans accounts for genome variability in a cosmopolitan phytoplankton. *ISME J.* **9**, 1365–1377 (2015).
99. Read, B. A. *et al.* Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* **499**, 209–213 (2013).
100. Karl, D. M., Church, M. J., Dore, J. E., Letelier, R. M. & Mahaffey, C. Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. *Proc. Natl Acad. Sci. USA* **109**, 1842–1849 (2012).
101. Benitez-Nelson, C. R. *et al.* Mesoscale eddies drive increased silica export in the subtropical Pacific Ocean. *Science* **316**, 1017–1021 (2007).
102. Lewin, J. C. Heterotrophy in diatoms. *J. Gen. Microbiol.* **9**, 305–313 (1953).
103. Croft, M. T., Warren, M. J. & Smith, A. G. Algae need their vitamins. *Eukaryot. Cell* **5**, 1175–1183 (2006).
104. Mitra, A. *et al.* Defining planktonic protist functional groups on mechanisms for energy and nutrient acquisition: incorporation of diverse mixotrophic strategies. *Protist* **167**, 106–120 (2016).
105. Ward, B. A. & Follows, M. J. Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux. *Proc. Natl Acad. Sci. USA* **113**, 2958–2963 (2016). **A recent attempt to introduce mixotrophic behaviour into global marine biogeochemical models.**
106. Liu, Z., Campbell, V., Heidelberg, K. B. & Caron, D. A. Gene expression characterizes different nutritional strategies among three mixotrophic protists. *FEMS Microbiol. Ecol.* **92**, fiv106 (2016).
107. Sanders, R. W., Caron, D. A., Davidson, J. M., Dennett, M. R. & Moran, D. M. Nutrient acquisition and population growth of a mixotrophic alga in axenic and bacterized cultures. *Microb. Ecol.* **42**, 513–523 (2001).
108. Stoecker, D. K. Mixotrophy among dinoflagellates. *J. Eukaryot. Microbiol.* **46**, 397–401 (1999).
109. Johnson, M. D. Acquired phototrophy in ciliates: a review of cellular interactions and structural adaptations. *J. Eukaryot. Microbiol.* **58**, 185–195 (2011).
110. Johnson, M. D., Oldach, D., Delwiche, D. F. & Stoecker, D. K. Retention of transcriptionally active cryptophyte nuclei by the ciliate *Myrionecta rubra*. *Nature* **445**, 426–428 (2007).
111. Lasek-Nesselquist, E., Wisecaver, J. H., Hackett, J. D. & Johnson, M. D. Insights into transcriptional changes that accompany organelle sequestration from the stolen nucleus of *Mesodinium rubrum*. *BMC Genomics* **16**, 805 (2015).
112. Lima-Mendez, G. *et al.* Determinants of community structure in the global plankton interactome. *Science* **348**, 1262073 (2015). **The introduction of the 'interactome', a concept for the organization of marine microbes into groups or 'guilds' of closely interacting taxa.**
113. Sillo, A. *et al.* Genome-wide transcriptional changes induced by phagocytosis or growth on bacteria in *Dictyostelium*. *BMC Genomics* **9**, 291 (2008).
114. Liu, Z. *et al.* Gene expression in the mixotrophic prymnesiophyte, *Prymnesium parvum*, responds to prey availability. *Front. Microbiol.* **6**, 319 (2015).
115. Barratt, J. L. N., Cao, M., Stark, D. J. & Ellis, J. T. The transcriptome sequence of *Dientamoeba fragilis* offers new biological insights on its metabolism, kinome, degradome and potential mechanisms of pathogenicity. *Protist* **166**, 389–408 (2015).
116. Lu, Y., Wohlrab, S., Glöckner, G., Guillou, L. & John, U. Genomic insights into processes driving the infection of *Alexandrium tamarense* by the parasitoid *Amoebophrya* sp. *Eukaryot. Cell* **13**, 1439–1449 (2014).
117. Lee, R. *et al.* Analysis of EST data of the marine protist *Oxyrrhis marina*, an emerging model for alveolate biology and evolution. *BMC Genomics* **15**, 122 (2014).
118. Guo, Z., Zhang, H. & Lin, S. Light-promoted rhodopsin expression and starvation survival in the marine dinoflagellate *Oxyrrhis marina*. *PLoS ONE* **9**, e114941 (2014).
119. Amin, S. A. *et al.* Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**, 98–101 (2015).
120. Stocker, R. Marine microbes see a sea of gradients. *Science* **338**, 628–633 (2012). **A report about the microscales of physics and chemistry, and their role in shaping aquatic microbial activities.**
121. Sapp, M. *et al.* Species-specific bacterial communities in the phycosphere of microalgae? *Microb. Ecol.* **53**, 683–699 (2007).
122. Schäfer, H., Abbas, B., Witte, H. & Muyzer, G. Genetic diversity of 'satellite' bacteria present in cultures of marine diatoms. *FEMS Microbiol. Ecol.* **42**, 25–35 (2002).
123. Durham, B. P. *et al.* Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc. Natl Acad. Sci. USA* **112**, 453–457 (2015).
124. Thompson, A. W. *et al.* Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science* **337**, 1546–1550 (2012).
125. Cruz-López, R. & Maske, H. The vitamin B<sub>1</sub> and B<sub>12</sub> required by the marine dinoflagellate *Lingulodinium polyedrum* can be provided by its associated bacterial community in culture. *Front. Microbiol.* **7**, 560 (2016).
126. Amin, S. A., Parker, M. S. & Armbrust, E. V. Interactions between diatoms and bacteria. *Microbiol. Mol. Biol. Rev.* **76**, 667–684 (2012).
127. Moustafa, A. *et al.* Transcriptome profiling of a toxic dinoflagellate reveals a gene-rich protist and a potential impact on gene expression due to bacterial presence. *PLoS ONE* **5**, e9688 (2010).
128. Choi, J. Y., Lee, T. W., Jeon, K. W. & Ahn, T. I. Evidence for symbiont-induced alteration of a host's gene expression: irreversible loss of SAM synthetase from *Amoeba proteus*. *J. Eukaryot. Microbiol.* **44**, 412–419 (1997).
129. Seyedsayamdost, M. R., Carr, G., Kolter, R. & Clardy, J. Roseobactin: small molecule modulators of an algal–bacterial symbiosis. *J. Am. Chem. Soc.* **133**, 18343–18349 (2011).
130. Woznica, A. *et al.* Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. *Proc. Natl Acad. Sci. USA* **113**, 7894–7899 (2016).
131. Van Etten, J. L., Lane, L. C. & Meints, R. H. Viruses and virulike particles of eukaryotic algae. *Microbiol. Rev.* **55**, 586–620 (1991).
132. Massana, R., Del Campo, J., Dinter, C. & Sommaruga, R. Crash of a population of the marine heterotrophic flagellate *Cafeteria roenbergensis* by viral infection. *Environ. Microbiol.* **9**, 2660–2669 (2007).
133. Hovde, B. T. *et al.* Genome sequence and transcriptome analyses of *Chrysochromulina tobin*: metabolic tools for enhanced algal fitness in the prominent Order Prymnesiales (Haptophyceae). *PLoS Genet.* **11**, e1005469 (2015).
134. Rokitta, S. D., von Dassow, P., Rost, B. & John, U. P- and N-depletion trigger similar cellular responses to promote senescence in eukaryotic phytoplankton. *Front. Mar. Sci.* **3**, 109 (2016).
135. Morgan-Smith, D., Clouse, M. A., Herndl, G. J. & Bochtansky, A. B. Diversity and distribution of microbial eukaryotes in the deep tropical and subtropical North Atlantic Ocean. *Deep Sea Res. Part I: Oceanogr. Res. Pap.* **78**, 58–69 (2013).
136. Farnelid, H. M., Turk-Kubo, K. A. & Zehr, J. P. Identification of associations between bacterioplankton and photosynthetic picoeukaryotes in coastal waters. *Front. Microbiol.* **7**, 339 (2016).
137. Yoon, H. S. *et al.* Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* **332**, 714–717 (2011). **The application of single-cell genomes is beginning to reveal the relationships among microbes.**
138. Veluchamy, A. *et al.* Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricornutum*. *Nat. Commun.* **4**, 2091 (2013).
139. Lovén, J. *et al.* Revisiting global gene expression analysis. *Cell* **151**, 476–482 (2012).
140. Nunn, B. L. *et al.* Diatom proteomics reveals unique acclimation strategies to mitigate Fe limitation. *PLoS ONE* **8**, e75653 (2013).
141. Poulson-Ellestad, K. L. *et al.* Metabolomics and proteomics reveal impacts of chemically mediated competition on marine plankton. *Proc. Natl Acad. Sci. USA* **111**, 9009–9014 (2014).
142. Aksoy, S. & Initiative, I. G. G. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of african trypanosomiasis. *Science* **344**, 380–386 (2014).
143. Fomes Flaten, G. A. *et al.* Studies of the microbial P-cycle during a Lagrangian phosphate-addition experiment in the Eastern Mediterranean. *Deep Sea Res. Part 2: Top. Stud. Oceanogr.* **52**, 2928–2943 (2005).
144. Strong, A., Chisholm, S., Miller, C. & Cullen, J. Ocean fertilization: time to move on. *Nature* **461**, 347–348 (2009).
145. Tekle, Y. I., Wegener Parfrey, L. & Katz, L. A. Molecular data are transforming hypotheses on the origin and diversification of eukaryotes. *Bioscience* **59**, 471–481 (2009).
146. Woese, C. R. Interpreting the universal phylogenetic tree. *Proc. Natl Acad. Sci. USA* **97**, 8392–8396 (2000).
147. Decelle, J., Colin, S. & Forster, R. A. in *Marine Protists: Diversity and Dynamics* (eds Ohtsuka, S., Suzuki, T., Horiguchi, T., Suzuki, N. & Not, F.) 465–500 (Springer, 2015).

#### Acknowledgements

The authors are grateful to the staff of the National Center for Genome Resources (NCGR; Santa Fe, New Mexico, USA) who assisted with sequencing (R. Bharti, J. Jacobi, J. Martinez, S. Nelson, P. Ngam and P. Umale), bioinformatics (C. Cameron, J. Crow, R. Kramer and K. Schilling) and software (N. Miller and K. Seal). They thank V. Chandler for input and guidance on the MMETSP and A. Lie for assistance with the figures. This research was funded in part by the Gordon and Betty Moore Foundation, through grants GBMF2637 to the NCGR and GBMF3111 to the National Center for Marine Algae and Microbiota (NCMA; East Boothbay, Maine, USA). Preparation of the manuscript was supported in part by a grant from the Simons Foundation (grant P49802 to D.A.C.).

#### Competing interests statement

The authors declare no competing interests.

#### FURTHER INFORMATION

**Gordon and Betty Moore Foundation MMETSP page:**  
<https://www.moore.org/grants/list/GBMF2637>  
**iMicrobe MMETSP:** <http://data.imicrobe.us/project/view/104>  
**MMETSP re-assemblies:**  
<https://monsterbashseq.wordpress.com>  
**NCBI BioProject:** <http://www.ncbi.nlm.nih.gov/bioproject/231566>

#### SUPPLEMENTARY INFORMATION

See online article: [S1](#) (box)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF