Eukaryotic plankton diversity in the sunlit ocean

Colomban de Vargas,[1,2]*† Stéphane Audic,[1,2]† Nicolas Henry,[1,2]† Johan Decelle,[1,2]†
Frédéric Mahé,[3,1,2]† Ramiro Logares,[4] Enrique Lara,[5] Cédric Berney,[1,2] Noan Le Bescot,[1,2]
Ian Probert,[6,7] Margaux Carmichael,[1,2,8] Julie Poulain,[9] Sarah Romac,[1,2]
Sébastien Colin,[1,2,8] Jean-Marc Aury,[9] Lucie Bittner,[10,11,8,1,2] Samuel Chaffron,[12,13,14]
Micah Dunthorn,[3] Stefan Engelen,[9] Olga Flegontova,[15,16] Lionel Guidi,[17,18] Aleš Horák,[15,16]
Olivier Jaillon,[9,19,20] Gipsi Lima-Mendez,[12,13,14] Julius Lukeš,[15,16,21] Shruti Malviya,[8]
Raphael Morard,[22,1,2] Matthieu Mulot,[5] Eleonora Scalco,[23] Raffaele Siano,[24]
Flora Vincent,[13,8] Adriana Zingone,[23] Céline Dimier,[1,2,8] Marc Picheral,[17,18]
Sarah Searson,[17,18] Stefanie Kandels-Lewis,[25,26] *Tara* Oceans Coordinators‡
Silvia G. Acinas,[4] Peer Bork,[25,27] Chris Bowler,[8] Gabriel Gorsky,[17,18] Nigel Grimsley,[28,29]
Pascal Hingamp,[30] Daniele Iudicone,[23] Fabrice Not,[1,2] Hiroyuki Ogata,[31]
Stephane Pesant,[32,22] Jeroen Raes,[12,13,14] Michael E. Sieracki,[33,34] Sabrina Speich,[35,36]
Lars Stemmann,[17,18] Shinichi Sunagawa,[25] Jean Weissenbach,[9,19,20]
Patrick Wincker,[9,19,20]* Eric Karsenti[26,8]*

Marine plankton support global biological and geochemical processes. Surveys of their biodiversity have hitherto been geographically restricted and have not accounted for the full range of plankton size. We assessed eukaryotic diversity from 334 size-fractionated photic-zone plankton communities collected across tropical and temperate oceans during the circumglobal *Tara* Oceans expedition. We analyzed 18S ribosomal DNA sequences across the intermediate plankton-size spectrum from the smallest unicellular eukaryotes (protists, >0.8 micrometers) to small animals of a few millimeters. Eukaryotic ribosomal diversity saturated at ~150,000 operational taxonomic units, about one-third of which could not be assigned to known eukaryotic groups. Diversity emerged at all taxonomic levels, both within the groups comprising the ~11,200 cataloged morphospecies of eukaryotic plankton and among twice as many other deep-branching lineages of unappreciated importance in plankton ecology studies. Most eukaryotic plankton biodiversity belonged to heterotrophic protistan groups, particularly those known to be parasites or symbiotic hosts.

The sunlit surface layer of the world's oceans functions as a giant biogeochemical membrane between the atmosphere and the ocean interior (*1*). This biome includes plankton communities that fix CO_2 and other elements into biological matter, which then enters the food web. This biological matter can be remineralized or exported to the deeper ocean, where it may be sequestered over ecological to geological time scales. Studies of this biome have typically focused on either conspicuous phyto- or zooplankton at the larger end of the organismal size spectrum or microbes (prokaryotes and viruses) at the smaller end. In this work, we studied the taxonomic and ecological diversity of the intermediate size spectrum (from 0.8 μm to a few millimeters), which includes all unicellular eukaryotes (protists) and ranges from the smallest protistan cells to small animals (*2*). The ecological biodiversity of marine planktonic protists has been analyzed using Sanger (*3–5*) and high-throughput (*6*, *7*) sequencing of mainly ribosomal DNA (rDNA) gene markers, on relatively small taxonomic and/or geographical scales, unveiling key new groups of phagotrophs (*8*), parasites (*9*), and phototrophs (*10*). We sequenced 18S rDNA metabarcodes up to local and global saturations from size-fractionated plankton communities sampled systematically across the world tropical and temperate sunlit oceans.

A global metabarcoding approach

To explore patterns of photic-zone eukaryotic plankton biodiversity, we generated ~766 million raw rDNA sequence reads from 334 plankton samples collected during the circumglobal *Tara* Oceans expedition (*11*). At each of 47 stations, plankton communities were sampled at two water-column depths corresponding to the main hydrographic structures of the photic zone: subsurface mixed-layer waters and the deep chlorophyll maximum (DCM) at the top of the thermocline. A low-shear, nonintrusive peristaltic pump and plankton nets of various mesh sizes were used on board *Tara* to sample and concentrate appropriate volumes of seawater to theoretically recover complete local eukaryotic biodiversity from four major organismal size fractions: piconanoplankton (0.8 to 5 μm), nanoplankton (5 to 20 μm), microplankton (20 to 180 μm), and mesoplankton (180 to 2000 μm) [see (*12*) for detailed *Tara* Oceans field sampling strategy and protocols].

We extracted total DNA from all samples, polymerase chain reaction (PCR)–amplified the hypervariable V9 region of the nuclear gene that encodes 18S rRNA (*13*), and generated an average of 1.73 ± 0.65 million sequence reads (paired-end Illumina) per sample (*11*). Strict bioinformatic quality control led to a final data set of 580 million reads, of which ~2.3 million were distinct,

hereafter denoted "metabarcodes." We then clustered metabarcodes into biologically meaningful operational taxonomic units (OTUs) (*14*) and assigned a eukaryotic taxonomic path to all metabarcodes and OTUs by global similarity analysis with 77,449 reference, Sanger-sequenced V9 rDNA barcodes covering the known diversity of eukaryotes and assembled into an in-house database called *V9_PR2* (*15*). Beyond taxonomic assignation, we inferred basic trophic and symbiotic ecological modes (photo- versus heterotrophy; parasitism, commensalism, mutualism for both hosts and symbionts) to *Tara* Oceans reads and OTUs on the basis of their genetic affiliation to large monophyletic and monofunctional groups of reference barcodes. We finally inferred large-scale ecological patterns of eukaryotic biodiversity across geography, taxonomy, and organismal size fractions based on rDNA abundance data and community similarity analyses and compared them to current knowledge extracted from the literature.

The extent of eukaryotic plankton diversity in the photic zone of the world ocean

Sequencing of ~1.7 million V9 rDNA reads from each of the 334 size-fractionated plankton samples was sufficient to approach saturation of eukaryotic richness at both local and global scales (Fig. 1, A and B). Local richness represented, on average, 9.7 ± 4% of global richness, the latter approaching saturation at ~2 million eukaryotic metabarcodes or ~110,000 OTUs (*16*). The global pool of OTUs displayed a good fit to the truncated Preston log-normal distribution (*17*), which, by extrapolation, suggests a total photic-zone eukaryotic plankton richness of ~150,000 OTUs, of which ~40,000 were not found in our survey (Fig. 1C). Thus, we estimate that our survey unveiled ~75% of eukaryotic ribosomal diversity in the globally distributed water masses analyzed. The extrapolated ~150,000 total OTUs is much higher than the ~11,200 formally described species of marine eukaryotic plankton (see below) and probably represents a highly conservative, lower-boundary estimate of the true number of eukaryotic species in this biome, given the relatively limited taxonomic resolution power of the 18*S* rDNA gene. Our data indicate that eukaryotic taxonomic diversity is higher in smaller organismal size fractions, with a peak in the piconanoplankton (Fig. 1A), highlighting the richness of tiny organisms that are poorly characterized in terms of morphotaxonomy and physiology (*18*). A first-order, supergroup-level classification of all *Tara* Oceans OTUs demonstrated the prevalence (at the biome scale and across the >four orders of size magnitude sampled) of protist rDNA biodiversity with respect to that of classical multicellular eukaryotes, i.e., animals, plants, and fungi (Fig. 2A). Protists accounted for >85% of total eukaryotic ribosomal diversity, a ratio that may well hold true for other marine, freshwater, and terrestrial oxygenic ecosystems (*19*). The latest estimates of total marine eukaryotic biodiversity based on statistical extrapolations from classical taxonomic knowledge predict the existence of 0.5 to 2.2 million species [including all benthic and planktonic systems from reefs to deep-sea vents (*20*, *21*)] but do not take into account the protistan knowledge gap highlighted here. Simple application of our animal–to–other eukaryotes ratio of ~13% to the robust prediction of the total number of metazoan species from (*20*) would imply that 16.5 million and 60 million eukaryotic species potentially inhabit the oceans and Earth, respectively.

Phylogenetic breakdown of photic-zone eukaryotic biodiversity

About one-third of eukaryotic ribosomal diversity in our data set did not match any reference barcode in the extensive *V9_PR2* database ("unassigned" category in Fig. 2A). This unassignable diversity represented only a small proportion (2.6%) of total reads and increased in both richness and abundance in smaller organismal size fractions, suggesting that it corresponds mostly to rare and minute taxa that have escaped previous characterization. Some may also correspond to divergent rDNA pseudogenes, known to exist in eukaryotes (*22*, *23*) or sequencing artefacts (*24*), although both of these would be expected to be present in equal proportion in all


Fig. 1. Photic-zone eukaryotic plankton ribosomal diversity. (**A**) V9 rDNA OTUs rarefaction curves and overall diversity (Shannon index, inset) for each plankton organismal size fraction. Proximity to saturation is indicated by weak slopes at the end of each rarefaction curve (e.g., 1.2/100,000 means 1.2 novel metabarcodes obtained every 100,000 rDNA reads sequenced). (**B**) Saturation slope versus number of V9 rDNA reads for all of the 334 samples (dots) analyzed herein. A slope of 0.02 indicates that two novel barcodes can be recovered if 100 new reads are sequenced. Samples are colored according to size fraction. (**C**) Global OTU abundance distribution and fit to the Preston log-normal model. Most OTUs in our data set were represented by 3 to 16 reads, whereas fewer OTUs presented less or more abundances. Quasi-Poisson fit to octaves (red curve) and maximized likelihood to log$_2$ abundances (blue curve) approximations were used to fit the OTU abundance distribution to the Preston log-normal model. Overall, the global (A) and local (B) saturation values indicate that our extensive sampling effort (in terms of spatiotemporal coverage and sequencing depth) uncovered the majority of eukaryotic ribosomal diversity within the photic layer of the world's tropical to temperate oceans. Calculation of the Preston veil, which infers the number of OTUs that we missed (or were veiled) during our sampling (~40,000), confirmed that we captured most of the protistan richness, thus allowing extraction of holistic and general patterns of eukaryotic plankton biodiversity from our data set.

size fractions [details in (16)]. The remaining ~87,000 assignable OTUs were classified into 97 deep-branching lineages covering the full spectrum of cataloged eukaryotic diversity amongst the seven recognized supergroups and multiple lineages of uncertain placement (15) whose origins go back to the primary radiation of eukaryotic life in the Neoproterozoic. Although highly represented in the *V9_PR2* reference database, several well-known lineages adapted to terrestrial, marine benthic, or anaerobic habitats (e.g., Embryophyta; apicomplexan and trypanosome parasites of land plants and animals; amoeboflagellate Breviatea; and several lineages of Amoebozoa, Excavata, and Cercozoa) were not detected in our metabarcoding data set, suggesting the absence of contamination during the PCR and sequencing steps on land and reducing the number of deep branches of eukaryotic plankton to 85 (Fig. 3).

We then extracted the metabarcodes assigned to morphologically well-known planktonic eukaryotic taxa from our data set and compared them with the conventional, 150 year-old morphological view of marine eukaryotic plankton that includes ~11,200 cataloged species divided into three broad categories: ~4350 species of phytoplankton (microalgae), ~1350 species of protozooplankton (relatively large, often biomineralized, heterotrophic protists), and ~5500 species of metazooplankton (holoplanktonic animals) (25–27). A congruent picture of the distribution of morphogenetic diversity among and within these organismal categories emerged from our data set (Fig. 2B), but typically, three to eight times more rDNA OTUs were found than described morphospecies in the best-known lineages within these categories. This is within the range of the number of cryptic species typically detected in globally-distributed pelagic taxa using molecular data (28, 29). The general congruency between genetic and morphological data in the cataloged compartment of eukaryotic plankton suggests that the protocols used, from plankton sampling to DNA sequencing, recovered the known eukaryotic biodiversity without major qualitative or quantitative biases. However, OTUs related to morphologically described taxa represented only a minor part of the total eukaryotic plankton ribosomal and phylogenetic diversity. Overall, <1% of OTUs were strictly identical to reference sequences, and OTUs were, on average, only ~86% similar to any V9 reference sequence (Fig. 3F) (16). This shows that most photic-zone eukaryotic plankton V9 rDNA diversity had not been previously sequenced from cultured strains, single-cell isolates, or even environmental clone library surveys. The *Tara* Oceans metabarcode data set added considerable phylogenetic information to previous protistan rDNA knowledge, with an estimated mean tree-length increase of 453%, reaching >100% in 43 lineages (16). Even in the best-referenced groups such as the diatoms (1232 reference sequences) (Fig. 3B), we identified many new rDNA sequences, both within known groups and forming new clades (16).

Eleven "hyperdiverse" lineages each contained >1000 OTUs, together representing ~88 and
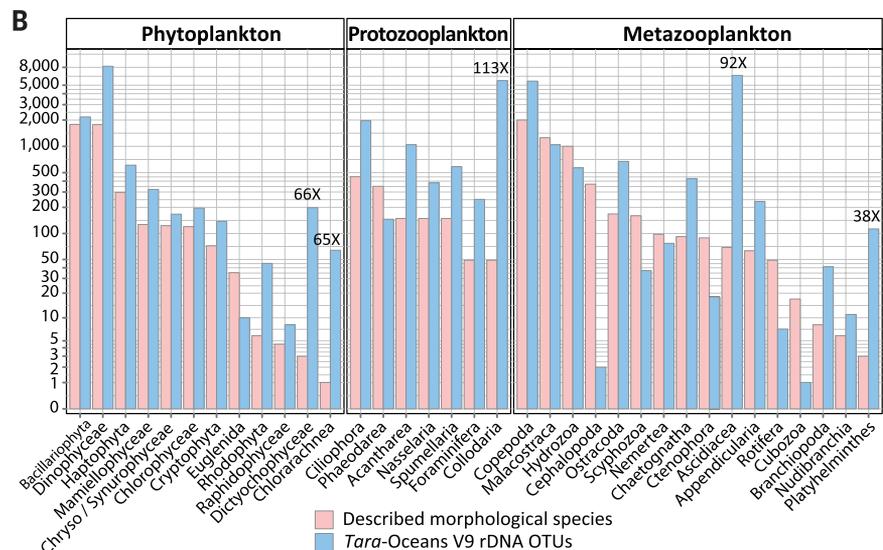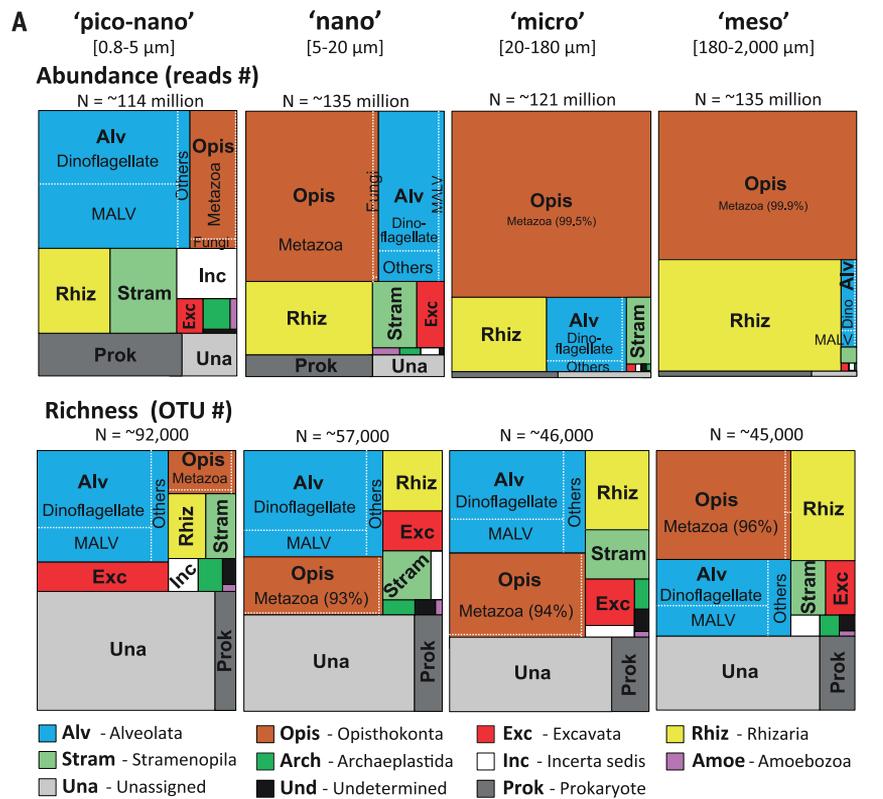
Fig. 2. Unknown and known components of eukaryotic plankton biodiversity. (A) Phylogenetic breakdown of the entire metabarcoding data set at the eukaryotic supergroup level. All *Tara* Oceans V9 rDNA reads and OTUs were classified among the seven recognized eukaryotic supergroups plus the known but unclassified deep-branching lineages (incertae sedis). The tree maps display the relative abundance (upper part) and richness (lower part) of the different eukaryotic supergroups in each organismal size fraction. Note that ~5% of barcodes were assigned to prokaryotes, essentially in the piconano fraction, witnessing the universality of the eukaryotic primers used. Barcodes are "unassigned" when sequence similarity to a reference sequence is <80% and "undetermined" when eukaryotic supergroups could not be discriminated (at similarity >80%). (B) Ribosomal DNA diversity associated with the morphologically known and cataloged part of eukaryotic plankton. The total number of morphologically described species in the literature [red bars, based on (25–27)] and the corresponding total number of *Tara* Oceans V9 rDNA OTUs (blue bars) are indicated for each of the 35 classical lineages of eukaryotic phyto-, protozoo-, and metazooplankton. The five classical groups that were found to be substantially more diverse than previously thought (from 38- to 113-fold more OTUs than morphospecies) are highlighted. Note that in the classical morphological view, phyto- and metazooplankton comprise ~88% of total eukaryotic plankton diversity.
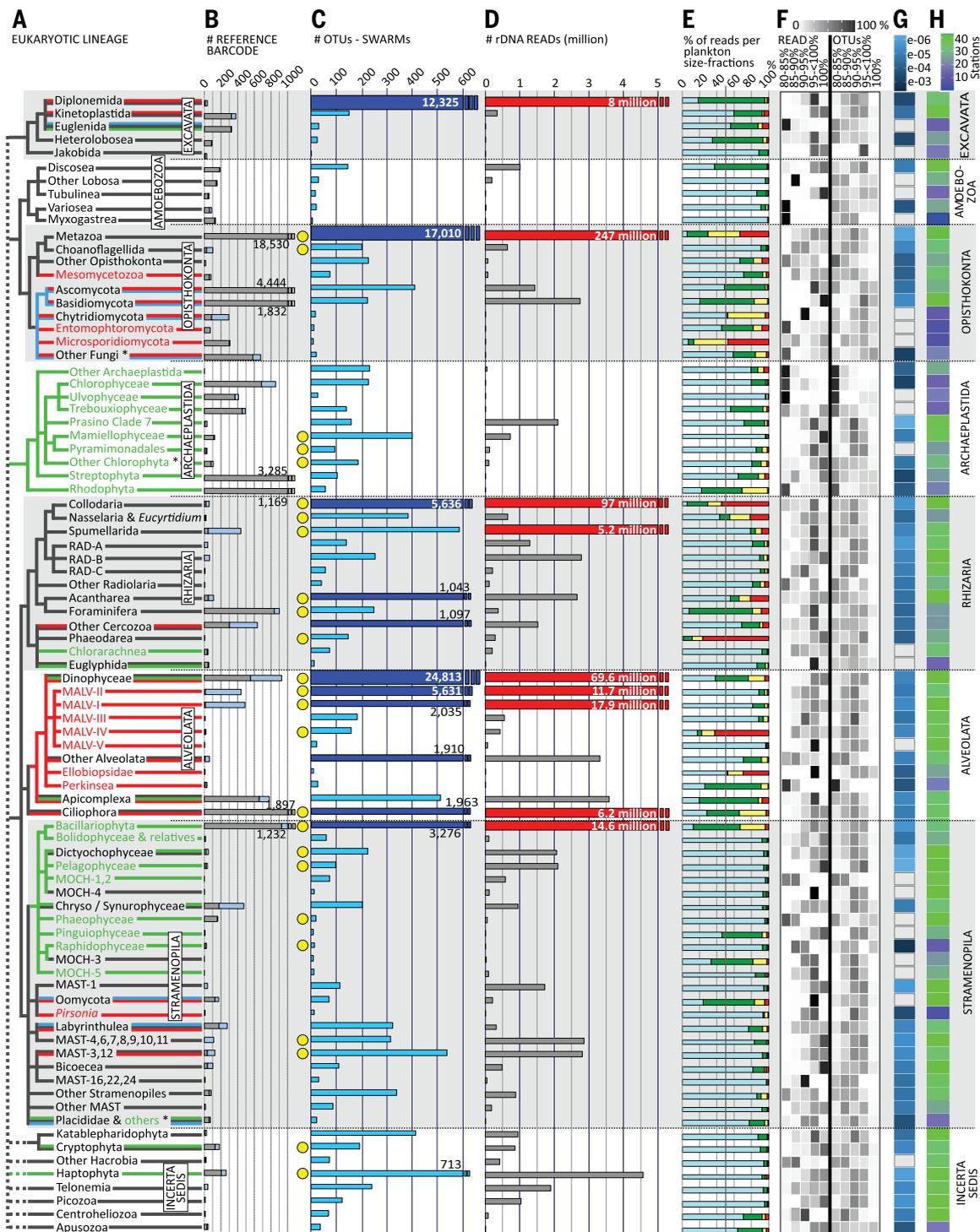
Fig. 3. Phylogenetic distribution of the assignable component of eukaryotic plankton ribosomal diversity. (**A**) Schematic phylogeny of the 85 deep-branching eukaryotic lineages represented in our global oceans metabarcoding data set, with broad ecological traits based on current knowledge: red, parasitic; green, photoautotrophic; blue, osmo- or saprotrophic; black, mostly phago-trophic lineages. Lineages known only from environmental sequence data were colored in black by default. For simplicity, three branches (denoted by asterisks) artificially group a few distinct lineages [details in (*15*)]. (**B**) Number of reference V9 rDNA barcodes used to annotate the metabarcoding data set (gray, with known taxonomy at the genus and/or species level; light blue, from previous 18S rDNA environmental clone libraries). (**C**) *Tara* Oceans V9 rDNA OTU richness.

Dark blue thicker bars indicate the 11 hyperdiverse lineages containing >1000 OTUs. Yellow circles highlight the 25 lineages that have been recognized as important in previous marine plankton biodiversity and ecology studies using morphological and/or molecular data [see also (*15*)]. (**D**) Eukaryotic plankton abundance expressed as numbers of rDNA reads (the red bars indicate the nine most abundant lineages with >5 million reads). (**E**) Proportion of rDNA reads per organismal size fraction. Light blue, piconano-; green, nano-; yellow, micro-; red, mesoplankton. (**F**) Percentage of reads and OTUs with 80 to 85%, 85 to 90%, 90 to 95%, 95 to <100%, and 100% sequence similarity to a reference sequence. (**G**) Slope of OTU rarefaction curves. (**H**) Mean geographic occupancy (average number of stations in which OTUs were observed, weighted by OTU abundance).

~90% of all OTUs and reads, respectively (Fig. 3C). Among these, the only permanently phototrophic taxa were diatoms (Fig. 4A) and about one-third of dinoflagellates (Fig. 4, B to F), together comprising ~15 and ~13% of hyperdiverse OTUs and reads, respectively (30). Most hyperdiverse photic-zone plankton belonged to three supergroups—the Alveolata, Rhizaria, and Excavata—about which we have limited biological or ecological information. The Alveolata, which consist mostly of parasitic [marine alveolates (MALVs)] (Fig. 4F) and phagotrophic (ciliates and most dinoflagellates) taxa, were by far the most diverse supergroup, comprising ~42% of all assignable OTUs. The Rhizaria are a group of amoeboid heterotrophic protists with active pseudopods displaying a broad spectrum of ecological behavior, from phagotrophy to parasitism and mutualism (symbioses) (31). Rhizarian diversity peaked in the Retaria (Fig. 4, C and D) a subgroup including giant protists that build complex skeletons of silicate (Polycystinea), strontium sulfate (Acantharia) (Fig. 4C), or calcium carbonate (Foraminifera) and thus comprise key microfossils for paleoceanography. Unsuspected rDNA diversity was recorded within the Collodaria (5636 OTUs), polycystines that are mostly colonial, poorly silicified, or naked and live in obligatory symbiosis with photosynthetic dinoflagellates (Fig. 4D) (32, 33). Arguably, the most surprising component of novel biodiversity was the >12,300 OTUs related to reference sequences of diplonemids, an excavate lineage that has only two described genera of flagellate grazers, one of which parasitizes diatoms and crustaceans (34, 35). Their ribosomal diversity was not only much higher than that observed in classical plankton groups such as foraminifers, ciliates, or diatoms (50-fold, 6-fold, and 3.8-fold higher, respectively) but was also far from richness saturation (Fig. 3E). Eukaryotic rDNA diversity peaked especially in the few lineages that extend across larger size fractions (i.e., metazoans, rhizarians, dinoflagellates, ciliates, diatoms) (Fig. 3E). Larger cells or colonies not only provide protection against predation via size-mediated avoidance and/or construction of composite skeletons but also provide support for complex and coevolving relationships with often specialized parasites or mutalistic symbionts.

Beyond this hyperdiverse, largely heterotrophic eukaryotic majority, our data set also highlighted the phylogenetic diversity of poorly known phagotrophic (e.g., 413 OTUs of Katablepharidophyta, 240 OTUs of Telonemia), osmotrophic (e.g., 410 OTUs of Ascomycota, 322 OTUs of Labyrinthulea), and parasitic (e.g., 384 OTUs of gregarine apicomplexans, 160 OTUs of Ascetosporea, 68
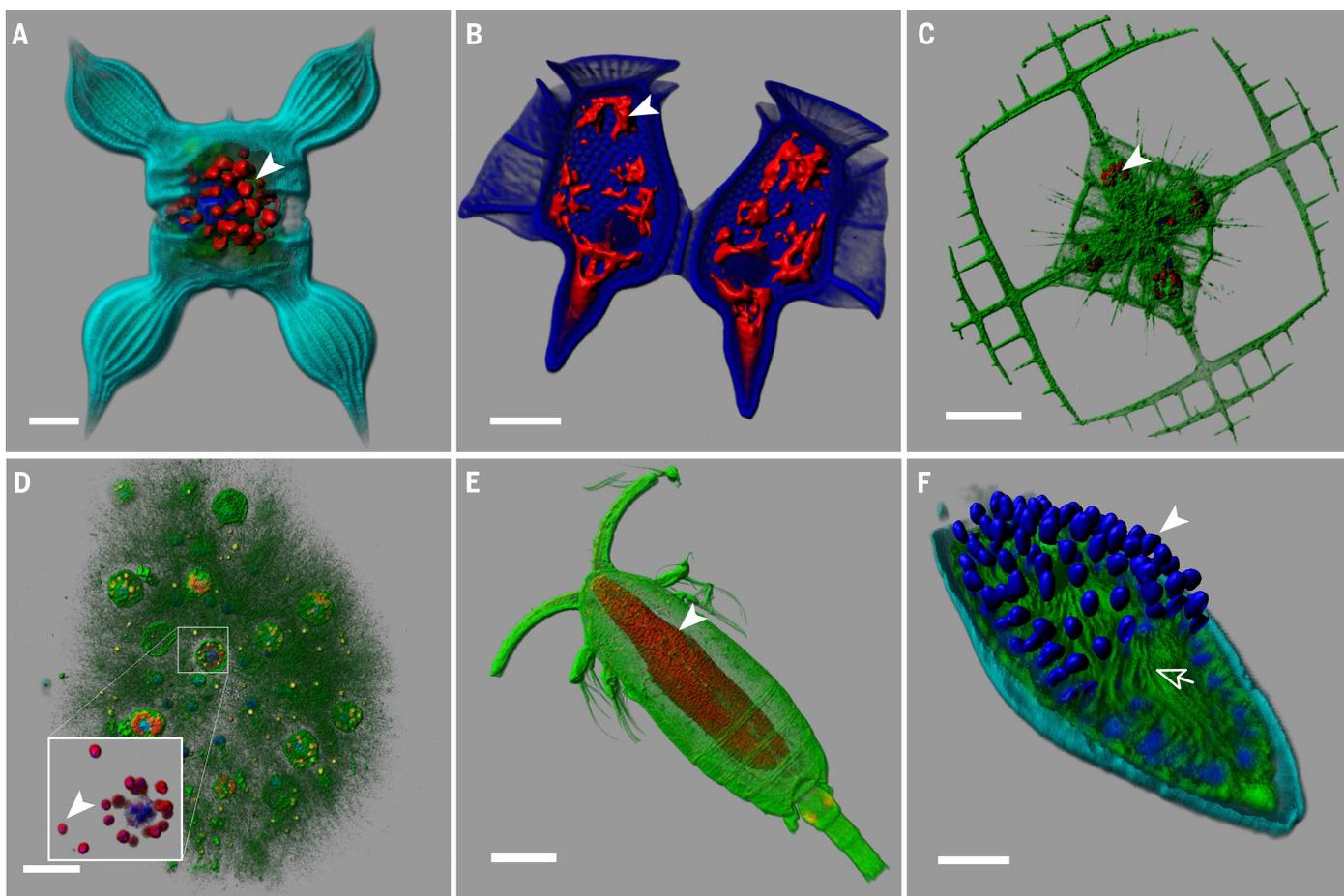
Fig. 4. Illustration of key eukaryotic plankton lineages. (**A**) Stramenopila; a phototrophic diatom *Chaetoceros bulbosus*, with its chloroplasts in red (arrowhead). Scale bar, 10 μm. (**B**) Alveolata; a heterotrophic dinoflagellate *Dinophysis caudata* harboring kleptoplasts [in red (arrowhead)]. Scale bar, 20 μm (75). (**C**) Rhizaria; an acantharian *Lithoptera* sp. with endosymbiotic haptophyte cells from the genus *Phaeocystis* [in red (arrowhead)]. Scale bar, 50 μm (41). (**D**) Rhizaria; inside a colonial network of Collodaria, a cell surrounded by several captive dinoflagellate symbionts of the genus *Brandtodinium* (arrowhead). Scale bar, 50 μm (33). (**E**) Opisthokonta; a copepod whose gut is colonized by the parasitic dinoflagellate *Blastodinium* [red area shows nuclei (arrowhead)]. Scale bar, 100 μm (51). (**F**) Alveolata; a cross-sectioned, dinoflagellate cell infected by the parasitoid alveolate *Amoebophrya* (MALV-II). Each blue spot (arrowhead) is the nucleus of future free-living dinospores; their flagella are visible in green inside the mastigocoel cavity (arrow). Scale bar, 5 μm. The cellular membranes were stained with DiOC6 (green); DNA and nuclei were stained with Hoechst (blue) [the dinoflagellate theca in (B) was also stained by this dye]. Chlorophyll autofluorescence is shown in red [except for in (E)]. An unspecific fluorescent painting of the cell surface (light blue) was used to reveal cell shape for (A) and (F). All specimens come from *Tara* Oceans samples preserved for confocal laser scanning fluorescent microscopy. Images were three-dimensionally reconstructed with Imaris (Bitplane).

OTUs of Ichthyosporea) protist groups. Amongst the 85 major lineages presented in the phylogenetic framework of Fig. 3, less than one-third (~25) have been recognized as important in previous marine plankton biodiversity and ecology studies using morphological and/or molecular data (Fig. 3C) (*15*). The remaining ~60 branches had either never been observed in marine plankton or were detected through morphological description of one or a few species and/or the presence of environmental sequences in geographically restricted clone library surveys (*15*). This understudied diversity represents ~25% of all taxonomically assignable OTUs (>21,500) and covers broad taxonomic and geographic scales, thus representing a wealth of new actors to integrate into future plankton systems biology studies.

Insights into photic-zone eukaryotic plankton ecology

Functional annotation of taxonomically assigned V9 rDNA metabarcodes was used as a first attempt to explore ecological patterns of eukaryotic diversity across broad spatial scales and organismal size fractions, focusing on fundamental trophic modes (photo- versus heterotrophy) and symbiotic interactions (parasitism to mutualism). Heterotroph (protists and metazoans) V9 rDNA metabarcodes were substantially more diverse (63%) and abundant (62%) than phototroph metabarcodes that represented <20% of OTUs and reads across all size fractions and geographic sites, with an increasing heterotroph-to-phototroph ratio in the micro- and mesoplankton (Fig. 5A, confirmed in 17 non–size-fractionated samples (*30*)). These results challenge the classical morphological view of plankton diversity, biased by a terrestrial ecology approach, whereby phyto- and metazooplankton (the plant-animal paradigm) are thought to comprise ~88% of eukaryotic plankton diversity (Fig. 2B) and heterotrophic protists are typically reduced in food-web modeling to a single entity, often idealized as ciliate grazers.

An unsuspected richness and abundance of metabarcodes assigned to monophyletic groups of heterotrophic protists that cannot survive without endosymbiotic microalgae was found in larger size fractions ("photosymbiotic hosts" in Fig. 5A). Their abundance and even diversity were sometimes greater than those of all metazoan metabarcodes, including those from copepods. Most of these cosmopolitan photosymbiotic hosts were found in the hyperdiverse radiolarians Acantharia (1043 OTUs) and Collodaria (5636 OTUs) (Figs. 3, 4B, and 5D), which have often been overlooked in traditional morphological surveys of plankton-net–collected material because of their delicate gelatinous and/or easily dissolved structures but are known to be very abundant from microscope-based and in situ imaging studies (*36–38*). All 95 known colonial collodarian species described since the 19th century (*39*) harbor intracellular symbiotic microalgae, and these key players for plankton ecology are protist analogs of photosymbiotic corals in

tropical coastal reef ecosystems with no equivalent in terrestrial ecology. In addition to their contribution to total primary production (*36*, *38*), these diverse, biologically complex, often biomineralized, and relatively long-lived giant mixotrophic protists stabilize carbon in larger size fractions and probably increase its flux to the ocean interior (*38*). Conversely, the microalgae that are known obligate intracellular partners in open-ocean photosymbioses (*33*, *40–42*) (Fig. 5B) were neither very diverse nor highly abundant and occurred evenly across organismal size fractions (Fig. 5C). However, their relative contribution was greatest in the mesoplankton category (10%) (Fig. 5C), where the known photosymbionts of pelagic rhizarians were found (together with their hosts) (Fig. 5B). The stable and systematic abundance of photosymbiotic microalgae across size fractions [a pattern not shown by nonphotosymbiotic microalgae (*30*)] suggests that pelagic photosymbionts maintain free-living and potentially actively growing populations in the piconano- and nanoplankton, representing an accessible pool for recruitment by their heterotrophic hosts. This appears to contrast with photosymbioses in coral reefs and terrestrial systems, where symbiotic microalgal populations mainly occur within their multicellular hosts (*43*).

On the other end of the spectrum of biological interactions, rDNA metabarcodes affiliated to groups of known parasites were ~90 times more diverse than photosymbionts in the piconanoplankton, where they represented ~59% of total heterotrophic protistan ribosomal richness and ~53% of abundance (Figs. 4 and 5C), although this latter value may be inflated by a hypothetically higher rDNA copy number in some marine alveolate lineages (*18*). Parasites in this size fraction were mostly (89% of diversity and 88% of abundance across all stations) within the MALV-I and -II Syndiniales (*30*), which are known exclusively as parasitoid species that kill their hosts and release hundreds of small (2 to 10 μm), nonphagotrophic dinospores (*9*, *44*) that survive for only a few days in the water column (*45*). Abundant parasite-assigned metabarcodes in small size fractions (Fig. 5, B and C) suggest the existence of a large and diverse pool of free-living parasites in photic-zone piconanoplankton, mirroring phage ecology (*46*) and reflecting the extreme diversity and abundance of their known main hosts: radiolarians, ciliates, and dinoflagellates (Fig. 3) (*9*, *47–49*). Contrasting with the pattern observed for metabarcodes affiliated to purely phagotrophic taxa, the relative abundance and richness of putative parasite metabarcodes decreased in the nano- and microplanktonic size fractions but increased again in the mesoplankton (Fig. 5C), where parasites are most likely in their infectious stage within larger-sized host organisms. This putative in hospite parasites richness, equivalent to only 23% of that in the piconanoplankton, consisted mostly of a variety of alveolate taxa known to infect crustaceans: MALV-IV such as *Haematodinium* and *Syndinium*; dinoflagellates such as *Blastodinium* (Fig. 4E); and apicomplexan gregarines, mainly *Cephaloidopho-*

roidea (Fig. 5B) (*9*, *50*, *51*). This pattern contrasts with terrestrial systems where most parasites live within their hosts and are typically transmitted either vertically or through vectors because they generally do not survive outside their hosts (*52*). In the pelagic realm, free-living parasitic spores, like phages, are protected from dessication and dispersed by water diffusion and are apparently massively produced, which likely increases horizontal transmission rate.

Community structuring of photic-zone eukaryotic plankton

Clustering of communities by their compositional similarity revealed the primary influence of organism size ($P = 10^{-3}$, $r^2 = 0.73$) on community structuring, with piconanoplankton displaying stronger cohesiveness than larger organismal size fractions (Fig. 6A). Filtered size-fraction–specific communities separated by thousands of kilometers were more similar in composition than they were to communities from other size fractions at the same location. This was emphasized by the fact that ~36% of all OTUs were restricted to a single size category (*53*). Further analyses within each organismal size fraction indicated that geography plays a role in community structuring, with samples being partially structured according to basin of origin, a pattern that was stronger in larger organismal size fractions ($P = 0.001$ in all cases, $r^2 = 0.255$ for piconanoplankton, 0.371 for nanoplankton, 0.473 for microplankton, and 0.570 for mesoplankton) (Fig. 6B). Mantel correlograms comparing Bray-Curtis community similarity to geographic distances between all samples indicated significant positive correlations in all organismal size fractions over the first ~6000 km, the correlation breaking down at larger geographic distances (*54*). This positive correlation between community dissimilarity and geographic distance, expected under neutral biodiversity dynamics (*55*), challenges the classical niche model for photic-zone eukaryotic plankton biogeography (*56*). The significantly stronger community differentiation by ocean basin in larger organismal size fractions (Fig. 6B) suggests increasing dispersal limitation from piconano- to nano-, micro-, and mesoplankton. Thus, larger-sized eukaryotic plankton communities, containing the highest abundance and diversity of metazoans (Figs. 2A and 5B), were spatially more heterogeneous in terms of both taxonomic (Fig. 6) and functional (Fig. 5A) composition and abundance. The complex life cycle and behaviors of metazooplankton, including temporal reproductive and growth cycles and vertical migrations, together with putative rapid adaptive evolution processes to mesoscale oceanographic features (*57*), may explain the stronger geographic differentiation of mesoplanktonic communities. By contrast, eukaryotic communities in the piconanoplankton were richer (Fig. 1A) and more homogeneous in taxonomic composition (Fig. 6), representing a stable compartment across the world's oceans (*58*).

Even though protistan communities were diverse, the proportions of abundant (>1%) and
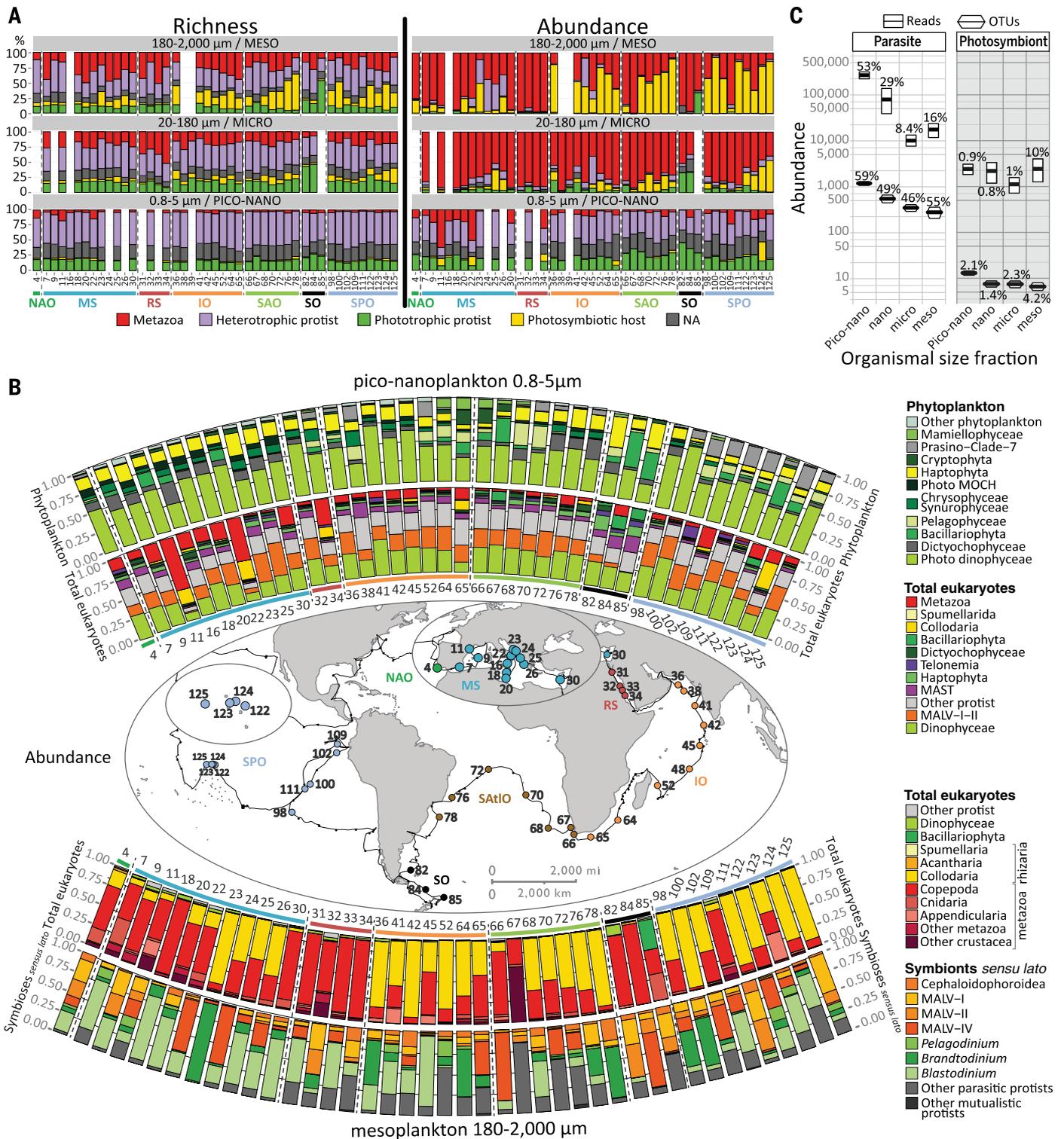
Fig. 5. Metabolizarcoding inference of trophic and symbiotic ecological diversity of photic-zone eukaryotic plankton. (**A**) Richness (OTU number) and abundance (read number) of rDNA metabarcodes assigned to various trophic taxo-groups across plankton organismal size fractions and stations. Note that the nano size fraction did not contain enough data to be used in this biogeographic analysis [for all size-fraction data, see (*30*)]. NA, not applicable. (**B**) Relative abundance of major eukaryotic taxa across *Tara* Oceans stations for (i) phytoplankton and all eukaryotes in piconanoplankton (above the map) and (ii) all eukaryotes and protistan symbionts (sensu

lato) in mesoplankton (below the map). Note the pattern of inverted relative abundance between collodarian colonies (Fig. 4) and copepods in, respectively, the oligotrophic and eutrophic and mesotrophic systems. The dinoflagellates *Brandtodinium* and *Pelagodinium* are endophotosymbionts in Collodaria (*33*) and Foraminifera (*40*, *42*), respectively. (**C**) Richness and abundance of parasitic and photosymbiotic (microalgae) protists across organismal size fractions. The relative contributions (percent) of parasites to total heterotrophic protists and of photosymbionts to total phytoplankton are indicated above each symbol.
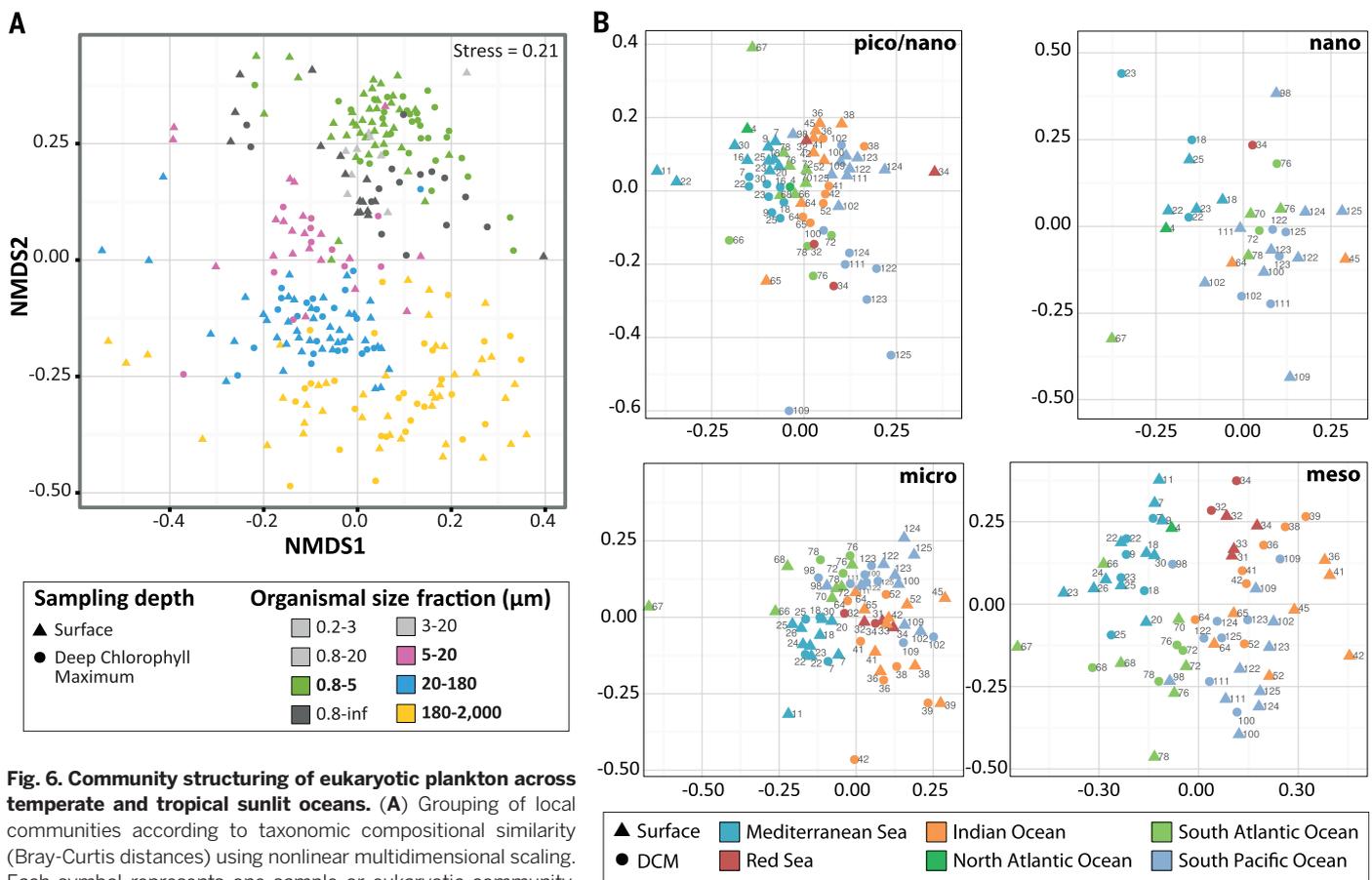
Fig. 6. Community structuring of eukaryotic plankton across temperate and tropical sunlit oceans. (**A**) Grouping of local communities according to taxonomic compositional similarity (Bray-Curtis distances) using nonlinear multidimensional scaling. Each symbol represents one sample or eukaryotic community, corresponding to a particular depth (shape) and organismal size fraction (color). (**B**) Same as in (A), but the different plankton organismal size fractions were analyzed independently, and communities are distinguished by depth (shape) and ocean basins' origin (color). An increasing geographic community differentiation along increasing organismal size fractions is visible and confirmed by the Mantel test [$P = 10^{-3}$, R_m = 0.36, 0.49, 0.50, and 0.51 for the highest piconano- to mesoplankton correlations in Mantel correlograms; see also (*54*)]. In addition, samples from the piconanoplankton only were discriminated by depth (surface versus DCM; P = 0.001, r^2 = 0.2). The higher diversity and abundance of eukaryotic phototrophs in this fraction (Fig. 5A) may explain overall community structuring by light and, thus, depth.

---

rare (<0.01%) OTUs were more or less constant across communities, as has been observed in coastal waters (*6*). Only 2 to 17 OTUs (i.e., 0.2 to 8% of total OTUs per and across sample) dominated each community (*54*), suggesting that a small proportion of eukaryotic taxa are key for local plankton ecosystem function. On a worldwide scale, an occurrence-versus-abundance analysis of all ~110,000 *Tara* Oceans OTUs revealed the hyperdominance of cosmopolitan taxa (Fig. 7A). The 381 (0.35% of the total) cosmopolitan OTUs represented ~68% of the total number of reads in the data set. Of these, 269 (71%) OTUs had >100,000 reads and accounted for nearly half (48%) of all rDNA reads (Fig. 7A), a pattern reminiscent of hyperdominance in the largest forest ecosystem on Earth, where only 227 tree species out of an estimated total of 16,000 account for half of all trees in Amazonia (*59*). The cosmopolitan OTUs belonged mainly (314 of 381) to the 11 hyperdiverse eukaryotic planktonic lineages (Fig. 3C) and were essentially phagotrophic (40%) or parasitic (21%), with relatively few (15%) phytoplanktonic taxa (*54*). Of the cosmopolitan OTUs, which represent organisms that are likely among the most abundant eukaryotes on Earth, 25% had poor identity (<95%) to reference taxa, and 11 of these OTUs could not even be affiliated to any available reference sequence (Fig. 7B) (*54*).

Conclusions and perspectives

We used rDNA sequence data to explore the taxonomic and ecological structure of total eukaryotic plankton from the photic oceanic biome, and we integrated these data with existing morphological knowledge. We found that eukaryotic plankton are more diverse than previously thought, especially heterotrophic protists, which may display a wide range of trophic modes (*60*) and include an unsuspected diversity of parasites and photosymbiotic taxa. Dominance of unicellular heterotrophs in plankton ecosystems likely emerged at the dawn of the radiation of eukaryotic cells, together with arguably their most important innovation: phagocytosis. The onset of eukaryophagy in the Neoproterozoic (*61*) probably led to adaptive radiation in heterotrophic eukaryotes through specialization of trophic modes and symbioses, opening novel serial biotic

ecological niches. The extensive codiversification of relatively large heterotrophic eukaryotes and their associated parasites supports the idea that biotic interactions, rather than competition for resources and space (*62*), are the primary forces driving organismal diversification in marine plankton systems. Based on rDNA, heterotrophic protists may be even more diverse than prokaryotes in the planktonic ecosystem (*63*). Given that organisms in highly diverse and abundant groups, such as the alveolates and rhizarians, can have genomes more complex than those of humans (*64*), eukaryotic plankton may contain a vast reservoir of unknown marine planktonic genes (*65*). Insights are developing into how heterotrophic protists contribute to a multilayered and integrated ecosystem. The protistan parasites and mutualistic symbionts increase connectivity and complexity of pelagic food webs (*66*, *67*) while contributing to the carbon quota of their larger, longer-lived, and often biomineralized symbiotic hosts, which themselves contribute to carbon export when they die. Decoding the ecological and evolutionary rules governing plankton diversity remains essential for understanding how the
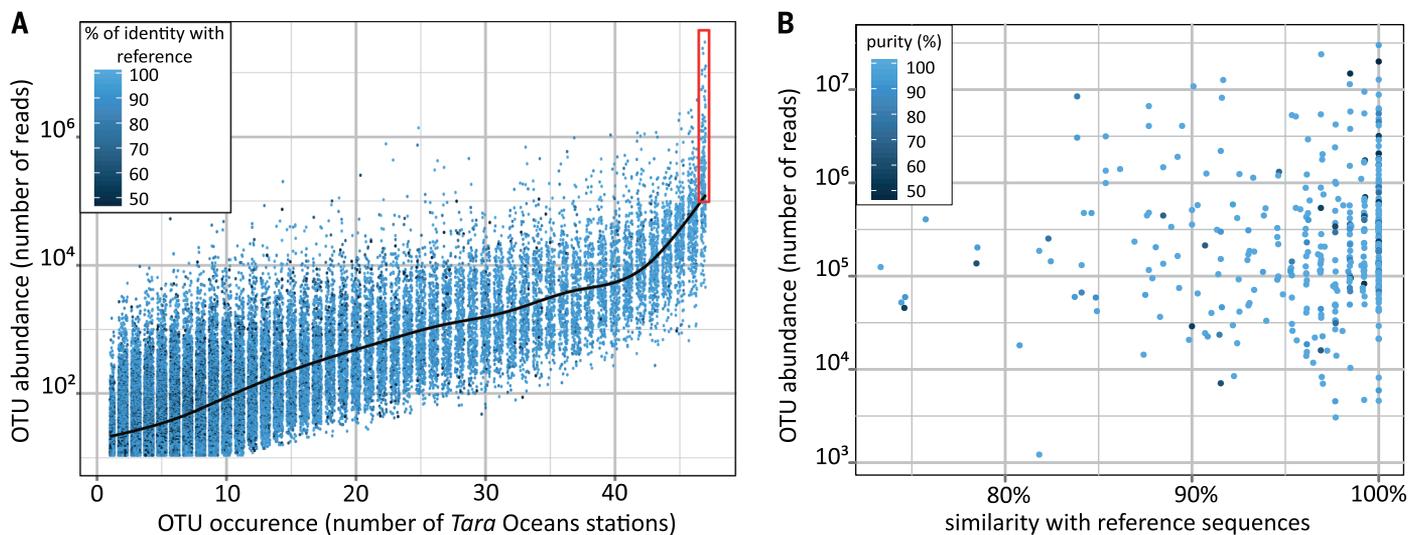
Fig. 7. Cosmopolitanism and abundance of eukaryotic marine plankton. (**A**) Occurrence-versus-abundance plot including the ~110,000 *Tara* Oceans V9 rDNA OTUs. OTUs are colored according to their identity with a reference sequence, and a fitted curve indicates the median OTU size value for each OTU geographic occurrence value. The red rectangle encloses the cosmopolitan and hyperdominant (>10^5 reads) OTUs. (**B**) Similarity to reference barcode and taxonomic purity [a measure of taxonomic assignment consistency defined as the percentage of reads within an OTU assigned to the same taxon; see (*13*)] of the 381 cosmopolitan OTUs, along their abundance (*y* axis).

critical ocean biomes contribute to the functioning of the Earth system.

Materials and methods

V9-18S rDNA for eukaryotic metabarcoding

We used universal eukaryotic primers (*68*) to PCR-amplify (25 cycles in triplicate) the V9-18*S* rDNA genes from all *Tara* Oceans samples. This barcode presents a combination of advantages for addressing general questions of eukaryotic biodiversity over extensive taxonomic and ecological scales: (i) It is universally conserved in length (130 ± 4 base pairs) and simple in secondary structure, thus allowing relatively unbiased PCR amplification across eukaryotic lineages followed by Illumina sequencing. (ii) It includes both stable and highly variable nucleotide positions over evolutionary time frames, allowing discrimination of taxa over a substantial phylogenetic depth. (iii) It is extensively represented in public reference databases across the eukaryotic tree of life, allowing taxonomic assignment among all known eukaryotic lineages (*13*).

Biodiversity analyses

Our bioinformatic pipeline included quality checking (Phred score filtering, elimination of reads without perfect forward and reverse primers, and chimera removal) and conservative filtering (removal of metabarcodes present in less than three reads and two distinct samples). The ~2.3 million metabarcodes (distinct reads) were clustered using an agglomerative, unsupervised single-linkage clustering algorithm, allowing OTUs to reach their natural limits while avoiding arbitrary global clustering thresholds (*13*, *14*). This clustering limited overestimation of biodiversity due to errors in PCR amplification or DNA sequencing, as well as intragenomic

polymorphism of rDNA gene copies (*13*). *Tara* Oceans metabarcodes and OTUs were taxonomically assigned by comparison to the 77,449 reference barcodes included in our *V9_PR2* database (*15*). This database derives from the Protist Ribosomal Reference (PR2) database (*69*) but focuses on the V9 region of the gene and includes the following reorganizations: (i) extension of the number of ranks for groups with finer taxonomy (e.g., animals), (ii) expert curation of the taxonomy and renaming in novel environmental groups and dinoflagellates, (iii) resolution of all taxonomic conflicts and inclusion of environmental sequences only if they provide additional phylogenetic information, and (iv) annotation of basic trophic and/or symbiotic modes for all reference barcodes assigned to the genus level [see (*53*) and (*15*) for details]. The *V9_PR2* reference barcodes represent 24,435 species and 13,432 genera from all known major lineages of the tree of eukaryotic life (*15*). Metabarcodes with ≥80% identity to a reference V9 rDNA barcode were considered assignable. Below this threshold it is not possible to discriminate between eukaryotic supergroups, given the short length of V9 rDNA sequences and the relatively fast rate accumulation of substitution mutations in the DNA. In addition to assignment at the finest-possible taxonomic resolution, all assignable metabarcodes were classified into a reference taxonomic framework consisting of 97 major monophyletic groups comprising all known high-rank eukaryotic diversity. This framework, primarily based on a synthesis of protistan biodiversity (*19*), also included all key but still unnamed planktonic clades revealed by previous environmental rDNA clone library surveys (*70*) [e.g., marine alveolates (MALV), marine stramenopiles (MAST), marine ochrophytes (MOCH), and radiolarians (RAD)] (*15*). Details of molecular and bioinformatics

methods are available on a companion Web site at http://taraoceans.sb-roscoff.fr/EukDiv/ (*53*). We compiled our data into two databases including the taxonomy, abundance, and size fraction and biogeography information associated with each metabarcode and OTU (*71*).

Ecological inferences

From our *Tara* Oceans metabarcoding data set, we inferred patterns of eukaryotic plankton functional ecology. Based on a literature survey, all reference barcodes assigned to at least the genus level that recruited *Tara* Oceans metabarcodes were associated to basic trophic and symbiotic modes of the organism they come from (*15*) and used for a taxo-functional annotation of our entire metabarcoding data set with the same set of rules used for taxonomic assignation (*53*). False positives were minimized by (i) assigning ecological modes to all individual reference barcodes in *V9_PR2*; (ii) inferring ecological modes to metabarcodes related to monomodal reference barcode(s) (otherwise transferring them to a "NA, nonapplicable" category); and (iii) exploring broad and complex trophic and symbiotic modes that involve fundamental reorganization of the cell structure and metabolism, emerged relatively rarely in the evolutionary history of eukaryotes, and most often concern all known species within monophyletic and ancient groups [see (*15*) for details]. In case of photo- versus heterotrophy, >75% of the major, deep-branching eukaryotic lineages considered (Fig. 3) are monomodal and recruit ~87 and ~69% of all *Tara* Oceans V9 rDNA reads and OTUs, respectively. For parasitism, ~91% of *Tara* Oceans metabarcodes are falling within monophyletic and major groups containing exclusively parasitic species (essentially within the major MALVs groups). Although biases could arise in functional annotation of metabarcodes

relatively distant from reference barcodes in the few complex polymodal groups (e.g., the dino-flagellates that can be phototrophic, heterotrophic, parasitic, or photosymbiotic), a conservative analysis of the trophic and symbiotic ecological patterns presented in Fig. 3, using a ≥99% assignation threshold, shows that these are stable across organismal size fractions and space, independently of the similarity cutoff (80 or 99%), demonstrating their robustness across evolutionary times (*30*).

Note that rDNA gene copy number varies from one to thousands in single eukaryotic genomes (*72*, *73*), precluding direct translation of rDNA read number into abundance of individual organisms. However, the number of rDNA copies per genome correlates positively to the size (*73*) and particularly to the biovolume (*72*) of the eukaryotic cell it represents. We compiled published data from the last ~20 years, confirming the positive correlation between eukaryotic cell size and rDNA copy number across a wide taxonomic and organismal size range [see (*74*); note, however, the ~one order of magnitude of cell size variation for a given rDNA copy number]. To verify whether our molecular ecology protocol preserved this empirical correlation, light microscopy counts of phytoplankton belonging to different eukaryotic supergroups (coccolithophores, diatoms, and dinoflagellates) were performed from nine *Tara* Oceans stations from the Indian, Atlantic, and Southern oceans; transformed into biomass and biovolume data; and then compared with the relative number of V9 rDNA reads found for the identified taxa in the same samples (*74*). Results confirmed the correlation between biovolume and V9 rDNA abundance data ($r^2 = 0.97$, $P = 1 \times 10^{-16}$), although we cannot rule out the possibility that some eukaryotic taxa may not follow the general trend.

REFERENCES AND NOTES

1. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998). doi: 10.1126/science.281.5374.237; pmid: 9657713

2. D. A. Caron, P. D. Countway, A. C. Jones, D. Y. Kim, A. Schnetzer, Marine protistan diversity. *Annu. Rev. Mar. Sci.* **4**, 467–493 (2012). doi: 10.1146/annurev-marine-120709-142802; pmid: 22457984

3. P. López-García, F. Rodríguez-Valera, C. Pedrós-Alió, D. Moreira, Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**, 603–607 (2001). doi: 10.1038/35054537; pmid: 11214316

4. S. Y. Moon-van der Staay, R. De Wachter, D. Vaulot, Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409**, 607–610 (2001). doi: 10.1038/35054541; pmid: 11214317

5. B. Díez, C. Pedrós-Alió, R. Massana, Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl. Environ. Microbiol.* **67**, 2932–2941 (2001). doi: 10.1128/AEM.67.7.2932-2941.2001; pmid: 11425705

6. R. Logares *et al.*, Patterns of rare and abundant marine microbial eukaryotes. *Curr. Biol.* **24**, 813–821 (2014). doi: 10.1016/j.cub.2014.02.050; pmid: 24704080

7. V. Edgcomb *et al.*, Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J.* **5**, 1344–1356 (2011). doi: 10.1038/ismej.2011.6; pmid: 21390079

8. R. Massana *et al.*, Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl. Environ. Microbiol.* **70**, 3528–3534 (2004). doi: 10.1128/AEM.70.6.3528-3534.2004; pmid: 15184153

9. L. Guillou *et al.*, Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environ. Microbiol.* **10**, 3349–3365 (2008). doi: 10.1111/j.1462-2920.2008.01731.x; pmid: 18771501

10. H. Liu *et al.*, Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12803–12808 (2009). doi: 10.1073/pnas.0905841106; pmid: 19622724

11. Companion Web site: Figure W1 and Database W1 (available at http://taraoceans.sb-roscoff.fr/EukDiv/).

12. S. Pesant *et al.*, Open science resources for the discovery and analysis of *Tara* Oceans data. http://biorxiv.org/content/early/2015/05/08/019117 (2015).

13. Companion Web site: Text W1 and Figure W2 (available at http://taraoceans.sb-roscoff.fr/EukDiv/).

14. F. Mahé, T. Rognes, C. Quince, C. de Vargas, M. Dunthorn, Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ* **2**, e593 (2014). doi: 10.7717/peerj.593; pmid: 25276506

15. Companion Web site: Database W2, Database W3, and Database W6 (available at http://taraoceans.sb-roscoff.fr/EukDiv/).

16. Companion Web site: Text W3, Text W4, Text W5, Figure W4, Figure W5, Figure W6, and Figure W7 (available at http://taraoceans.sb-roscoff.fr/EukDiv/).

17. F. W. Preston, The commonness, and rarity, of species. *Ecology* **29**, 254–283 (1948). doi: 10.2307/1930989

18. R. Massana, Eukaryotic picoplankton in surface oceans. *Annu. Rev. Microbiol.* **65**, 91–110 (2011). doi: 10.1146/annurev-micro-090110-102903; pmid: 21639789

19. J. Pawlowski *et al.*; CBOL protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLOS Biol.* **10**, e1001419 (2012). doi: 10.1371/journal.pbio.1001419; pmid: 23139639

20. C. Mora, D. P. Tittensor, S. Adl, A. G. B. Simpson, B. Worm, How many species are there on Earth and in the ocean? *PLOS Biol.* **9**, e1001127 (2011). doi: 10.1371/journal.pbio.1001127; pmid: 21886479

21. W. Appeltans *et al.*, The magnitude of global marine species diversity. *Curr. Biol.* **22**, 2189–2202 (2012). doi: 10.1016/j.cub.2012.09.036; pmid: 23159596

22. L. M. Márquez, D. J. Miller, J. B. MacKenzie, M. J. H. Van Oppen, Pseudogenes contribute to the extreme diversity of nuclear ribosomal DNA in the hard coral Acropora. *Mol. Biol. Evol.* **20**, 1077–1086 (2003). doi: 10.1093/molbev/msg122; pmid: 12777522

23. S. R. Santos, R. A. Kinzie III, K. Sakai, M. A. Coffroth, Molecular characterization of nuclear small subunit (18S)-rDNA pseudogenes in a symbiotic dinoflagellate (*Symbiodinium*, Dinophyta). *J. Eukaryot. Microbiol.* **50**, 417–421 (2003). doi: 10.1111/j.1550-7408.2003.tb00264.x; pmid: 14733432

24. J. Decelle, S. Romac, E. Sasaki, F. Not, F. Mahé, Intracellular diversity of the V4 and V9 regions of the 18S rRNA in marine protists (radiolarians) assessed by high-throughput sequencing. *PLOS ONE* **9**, e104297 (2014). doi: 10.1371/journal.pone.0104297; pmid: 25090095

25. A. Sournia, M.-J. Chrétiennot-Dinet, M. Ricard, Marine phytoplankton: How many species in the world ocean? *J. Plankton Res.* **13**, 1093–1099 (1991). doi: 10.1093/plankt/13.5.1093

26. P. H. Wiebe *et al.*, Deep-sea sampling on CMarZ cruises in the Atlantic Ocean – An introduction. *Deep-Sea Res. Part II* **57**, 2157–2166 (2010). doi: 10.1016/j.dsr2.2010.09.018

27. D. Boltovskoy, Diversity and endemism in cold waters of the South Atlantic: Contrasting patterns in the plankton and the benthos. *Sci. Mar.* **69**, 17–26 (2005).

28. C. de Vargas, R. Norris, L. Zaninetti, S. W. Gibb, J. Pawlowski, Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2864–2868 (1999). doi: 10.1073/pnas.96.6.2864; pmid: 10077602

29. K. M. K. Halbert, E. Goetze, D. B. Carlon, High cryptic diversity across the global range of the migratory planktonic copepods Pleuromamma piseki and P. gracilis. *PLOS ONE* **8**, e77011 (2013). doi: 10.1371/journal.pone.0077011; pmid: 24167556

30. Companion Web site: Figure W8, Figure W9, Figure W10, and Figure W14 (available at http://taraoceans.sb-roscoff.fr/EukDiv/).

31. F. Burki, P. J. Keeling, Rhizaria. *Curr. Biol.* **24**, R103–R107 (2014). doi: 10.1016/j.cub.2013.12.025; pmid: 24502779

32. N. R. Swanberg, thesis, Massachussetts Institute of Technology (1974).

33. I. Probert *et al.*, *Brandtodinium* gen. nov. and B. nutricula comb. Nov. (Dinophyceae), a dinoflagellate commonly found in symbiosis with polycystine radiolarians. *J. Phycol.* **50**, 388–399 (2014). doi: 10.1111/jpy.12174

34. S. von der Heyden, E. E. Chao, K. Vickerman, T. Cavalier-Smith, Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of euglenozoa. *J. Eukaryot. Microbiol.* **51**, 402–416 (2004). doi: 10.1111/j.1550-7408.2004.tb00387.x; pmid: 15352322

35. E. Schnepf, Light and electron microscopical observations in Rhynchopus coscinodiscivorus spec. nov., a Colorless, phagotrophic euglenozoon with concealed flagella. *Arch. Protistenkd.* **144**, 63–74 (1994). doi: 10.1016/S0003-9365(11)80225-3

36. M. R. Dennett, Video plankton recorder reveals high abundances of colonial Radiolaria in surface waters of the central North Pacific. *J. Plankton Res.* **24**, 797–805 (2002). doi: 10.1093/plankt/24.8.797

37. L. Stemmann *et al.*, Global zoogeography of fragile macrozooplankton in the upper 100-1000 m inferred from the underwater video profiler. *ICES J. Mar. Sci.* **65**, 433–442 (2008). doi: 10.1093/icesjms/fsn010

38. A. F. Michaels, D. A. Caron, N. R. Swanberg, F. A. Howse, C. M. Michaels, Planktonic sarcodines (Acantharia, Radiolaria, Foraminifera) in surface waters near Bermuda: Abundance, biomass and vertical flux. *J. Plankton Res.* **17**, 131–163 (1995). doi: 10.1093/plankt/17.1.131

39. E. Haeckel, "Report on the Radiolaria collected by H.M.S. Challenger during the years 1873-1876" in *Report on the Scientific Results of the Voyage of H.M.S. Challenger During the Years 1873-76. Zoology*. (Neill, Edinburgh, 1887).

40. R. Siano, M. Montresor, I. Probert, F. Not, C. de Vargas, *Pelagodinium* gen. nov. and P. béii comb. nov., a dinoflagellate symbiont of planktonic foraminifera. *Protist* **161**, 385–399 (2010). doi: 10.1016/j.protis.2010.01.002; pmid: 20149979

41. J. Decelle *et al.*, An original mode of symbiosis in open ocean plankton. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18000–18005 (2012). doi: 10.1073/pnas.1212303109; pmid: 23071304

42. Y. Shaked, C. de Vargas, Pelagic photosymbiosis: rDNA assessment of diversity and evolution of dinoflagellate symbionts and planktonic foraminiferal hosts. *Mar. Ecol. Prog. Ser.* **325**, 59–71 (2006). doi: 10.3354/meps325059

43. J. Decelle, New perspectives on the functioning and evolution of photosymbiosis in plankton: Mutualism or parasitism? *Commun. Integr. Biol.* **6**, e24560 (2013). doi: 10.4161/cib.24560; pmid: 23986805

44. R. Siano *et al.*, Distribution and host diversity of Amoebophryidae parasites across oligotrophic waters of the Mediterranean Sea. *Biogeosciences* **8**, 267–278 (2011). doi: 10.5194/bg-8-267-2011

45. D. Coats, M. Park, Parasitism of photosynthetic dinoflagellates by three strains of Amoebophrya (Dinophyta): Parasite survival, infectivity, generation time, and host specificity. *J. Phycol.* **528**, 520–528 (2002). doi: 10.1046/j.1529-8817.2002.01200.x

46. K. E. Wommack, R. R. Colwell, Virioplankton: Viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000). doi: 10.1128/MMBR.64.1.69-114.2000; pmid: 10704475

47. A. Skovgaard, Dirty tricks in the plankton: Diversity and role of marine parasitic protists. *Acta Protozool.* **53**, 51–62 (2014).

48. J. Bråte *et al.*, Radiolaria associated with large diversity of marine alveolates. *Protist* **163**, 767–777 (2012). doi: 10.1016/j.protis.2012.04.004; pmid: 22658831

49. T. R. Bachvaroff, S. Kim, L. Guillou, C. F. Delwiche, D. W. Coats, Molecular diversity of the syndinean genus Eudoboscquella based on single-cell PCR analysis. *Appl. Environ. Microbiol.* **78**, 334–345 (2012). doi: 10.1128/AEM.06678-11; pmid: 22081578

50. S. Rueckert, T. G. Simdyanov, V. V. Aleoshin, B. S. Leander, Identification of a divergent environmental DNA sequence clade using the phylogeny of gregarine parasites (Apicomplexa) from crustacean hosts. *PLOS ONE* **6**, e18163 (2011). doi: 10.1371/journal.pone.0018163; pmid: 21483868

51. A. Skovgaard, S. A. Karpov, L. Guillou, The parasitic dinoflagellates Blastodinium spp. inhabiting the gut of marine, planktonic copepods: Morphology, ecology, and unrecognized species diversity. *Front. Microbiol.* **3**, 305 (2012). doi: 10.3389/fmicb.2012.00305; pmid: 22973263

52. H. McCallum *et al.*, Does terrestrial epidemiology apply to marine systems? *Trends Ecol. Evol.* **19**, 585–591 (2004). doi: 10.1016/j.tree.2004.08.009

53. Companion Web site: detailed Material and Methods, Database W9, and Figure W11 (available at http://taraoceans.sb-roscoff.fr/EukDiv/).

54. Companion Web site: Figure W12, Figure W13, Database W7, and Database W8 (available at http://taraoceans.sb-roscoff.fr/EukDiv/).

55. M. Holyoak, M. A. Leibold, R. D. Holt, Eds., *Metacommunities: Spatial Dynamics and Ecological Communities* (University of Chicago Press, Chicago, 2005).

56. L. G. M. Baas Becking, *Geobiologie of Inleiding tot de Milieukunde* (W. P. Van Stockum and Zoon, The Hague, Netherlands, 1934).

57. K. T. C. A. Peijnenburg, E. Goetze, High evolutionary potential of marine zooplankton. *Ecol. Evol.* 3, 2765–2781 (2013). doi: 10.1002/ece3.644; pmid: 24567838

58. V. Smetacek, Microbial food webs. The ocean's veil. *Nature* 419, 565 (2002). doi: 10.1038/419565a; pmid: 12374956

59. H. ter Steege *et al.*, Hyperdominance in the Amazonian tree flora. *Science* 342, 1243092 (2013). doi: 10.1126/science.1243092; pmid: 24136971

60. D. Vaulot, K. Romari, F. Not, Are autotrophs less diverse than heterotrophs in marine picoplankton? 10, 266–267 (2002).

61. A. H. Knoll, Paleobiological perspectives on early eukaryotic evolution. *Cold Spring Harb. Perspect. Biol.* 6, 1–14 (2014). doi: 10.1101/cshperspect.a016121; pmid: 24384569

62. V. Smetacek, A watery arms race. *Nature* 411, 745 (2001). doi: 10.1038/35081210; pmid: 11459035

63. S. Sunagawa *et al.*, Structure and function of the global ocean microbiome. *Science* 348, 1261359 (2015).

64. M. J. Oliver, D. Petrov, D. Ackerly, P. Falkowski, O. M. Schofield, The mode and tempo of genome size evolution in eukaryotes. *Genome Res.* 17, 594–601 (2007). doi: 10.1101/gr.6096207; pmid: 17420184

65. H. Abida *et al.*, Bioprospecting marine plankton. *Mar. Drugs* 11, 4594–4611 (2013). doi: 10.3390/md11114594; pmid: 24240981

66. G. Lima-Mendez *et al.*, Determinants of community structure in the global plankton interactome. *Science* 348, 1262073 (2015).

67. K. D. Lafferty, A. P. Dobson, A. M. Kuris, Parasites dominate food web links. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11211–11216 (2006). doi: 10.1073/pnas.0604755103; pmid: 16844774

68. L. A. Amaral-Zettler, E. A. McCliment, H. W. Ducklow, S. M. Huse, A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLOS ONE* 4, e6372 (2009). doi: 10.1371/journal.pone.0006372; pmid: 19633714

69. L. Guillou *et al.*, The Protist Ribosomal Reference database (PR²): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* 41, D597–D604 (2013). doi: 10.1093/nar/gks1160; pmid: 23193267

70. R. Massana, J. del Campo, M. E. Sieracki, S. Audic, R. Logares, Exploring the uncultured microeukaryote majority in the oceans: Reevaluation of ribogroups within stramenopiles. *ISME J.* 8, 854–866 (2014). doi: 10.1038/ismej.2013.204; pmid: 24196325

71. Companion Web site: Database W4 and Database W5 (available at http://taraoceans.sb-roscoff.fr/EukDiv/).

72. A. Godhe *et al.*, Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. *Appl. Environ. Microbiol.* 74, 7174–7182 (2008). doi: 10.1128/AEM.01298-08; pmid: 18849462

73. F. Zhu, R. Massana, F. Not, D. Marie, D. Vaulot, Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol. Ecol.* 52, 79–92 (2005). doi: 10.1016/j.femsec.2004.10.006; pmid: 16329895

74. Companion Web site: Text W2 and Figure W3 (available at http://taraoceans.sb-roscoff.fr/EukDiv/).

75. M. Kim, S. Nam, W. Shin, D. W. Coats, M. Park, *Dinophysis caudata* (dinophyceae) sequesters and retains plastids from the mixotrophic ciliate prey *Mesodinium rubrum*. *J. Phycol.* 48, 569–579 (2012). doi: 10.1111/j.1529-8817.2012.01150.x

ACKNOWLEDGMENTS

Tara Oceans Coordinators

Silvia G. Acinas,[1] Peer Bork,[2] Emmanuel Boss,[3] Chris Bowler,[4] Colomban de Vargas,[5,6] Michael Follows,[7] Gabriel Gorsky,[8,9] Nigel Grimsley,[10,11] Pascal Hingamp,[12] Daniele Iudicone,[13] Olivier Jaillon,[14,15,16] Stefanie Kandels-Lewis,[2,17] Lee Karp-Boss,[3] Eric Karsenti,[4,17] Uros Krzic,[18] Fabrice Not,[5,6] Hiroyuki Ogata,[19] Stephane Pesant,[20,21] Jeroen Raes,[22,23,24] Emmanuel G. Reynaud,[25] Christian Sardet,[26,27] Mike Sieracki,[28] Sabrina Speich,[29,30] Lars Stemmann,[8] Matthew B. Sullivan,[31]* Shinichi Sunagawa,[2] Didier Velayoudon,[32] Jean Weissenbach,[14,15,16] Patrick Wincker[14,15,16]

[1]Department of Marine Biology and Oceanography, ICM-CSIC, Passeig Marítim de la Barceloneta, 37-49, Barcelona E08003, Spain. [2]Structural and Computational Biology, EMBL, Meyerhofstraße 1, 69117 Heidelberg, Germany. [3]School of Marine Sciences, University of Maine, Orono, ME 04469, USA. [4]Ecole Normale Supérieure, IBENS, and Inserm U1024, and CNRS UMR 8197, Paris, F-75005 France. [5]CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France. [6]Sorbonne Universités, UPMC Paris 06, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France. [7]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [8]CNRS, UMR 7093, LOV, Observatoire Océanologique, F-06230 Villefranche-sur-Mer, France. [9]Sorbonne Universités, UPMC Paris 06, UMR 7093, LOV, Observatoire Océanologique, F-06230 Villefranche-sur-Mer, France. [10]CNRS UMR 7232, BIOM, Avenue du Fontaulé, 66650 Banyuls-sur-Mer, France. [11]Sorbonne Universités Paris 06, OOB UPMC, Avenue du Fontaulé, 66650 Banyuls-sur-Mer, France. [12]Aix Marseille Université, CNRS IGS, UMR 7256, 13288 Marseille, France. [13]Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy. [14]CEA, Institut de Génomique, GENOSCOPE, 2 rue Gaston Crémieux, 91057 Evry, France. [15]CNRS, UMR 8030, CP5706 Evry, France. [16]Université d'Evry, UMR 8030, CP5706 Evry, France. [17]Directors' Research, EMBL, Meyerhofstraße 1, 69117 Heidelberg, Germany. [18]Cell Biology and Biophysics, EMBL, Meyerhofstraße 1, 69117 Heidelberg, Germany. [19]Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan. [20]PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, Bremen, Germany. [21]MARUM, Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany. [22]Department of Microbiology and Immunology, Rega Institute, KU Leuven, Herestraat 49, 3000 Leuven, Belgium. [23]Center for the Biology of Disease, VIB, Herestraat 49, 3000 Leuven, Belgium. [24]Department of Applied Biological Sciences, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. [25]Earth Institute, University College Dublin, Dublin, Ireland. [26]CNRS, UMR 7009 Biodev, Observatoire Océanologique, F-06230 Villefranche-sur-Mer, France. [27]Sorbonne Universités, UPMC Univ Paris 06, UMR 7009 Biodev, F-06230 Observatoire Océanologique, Villefranche-sur-Mer, France. [28]Bigelow Laboratory for Ocean Sciences, East Boothbay, ME 04544, USA. [29]Department of Geosciences, LMD, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris, Cedex 05, France. [30]Laboratoire de Physique des Océan UBO-IUEM Place Copernic 29820 Plouzané, France. [31]Department of Ecology and Evolutionary Biology, University of Arizona, 1007 East Lowell Street, Tucson, AZ 85721, USA. [32]DVIP Consulting, Sèvres, France.
*Present address: Department of Microbiology, Ohio State University, Columbus, OH 43210, USA.

SUPPLEMENTARY MATERIALS