



Tara Oceans: towards global ocean ecosystems biology

Shinichi Sunagawa¹✉, Silvia G. Acinas², Peer Bork^{3,4,5}, Chris Bowler^{6,7},
*Tara Oceans Coordinators**, Damien Eveillard¹, Gabriel Gorsky¹, Lionel Guidi¹,
 Daniele Iudicone¹⁰, Eric Karsenti^{6,7,11}, Fabien Lombard¹, Hiroyuki Ogata¹²,
 Stephane Pesant^{13,14}, Matthew B. Sullivan^{15,16,17}, Patrick Wincker¹ and
 Colomban de Vargas^{7,19}✉

Abstract | A planetary-scale understanding of the ocean ecosystem, particularly in light of climate change, is crucial. Here, we review the work of *Tara Oceans*, an international, multidisciplinary project to assess the complexity of ocean life across comprehensive taxonomic and spatial scales. Using a modified sailing boat, the team sampled plankton at 210 globally distributed sites at depths down to 1,000 m. We describe publicly available resources of molecular, morphological and environmental data, and discuss how an ecosystems biology approach has expanded our understanding of plankton diversity and ecology in the ocean as a planetary, interconnected ecosystem. These efforts illustrate how global-scale concepts and data can help to integrate biological complexity into models and serve as a baseline for assessing ecosystem changes and the future habitability of our planet in the Anthropocene epoch.

Epipelagic

Referring to uppermost layer of the ocean that receives sunlight, enabling the organisms inhabiting it to perform photosynthesis.

Mesopelagic

Referring to the ocean layer that receives very little to no sunlight, lying beneath the epipelagic layer, ranging from about 200 to 1,000 m in depth.

The *Tara Oceans* project

The ocean ecosystem covers ~70% of Earth's surface and contains 97% of all water on our planet. Plankton are the dominant life forms in the ocean and comprise highly dynamic and interacting populations of viruses, bacteria, archaea, single-celled eukaryotes (protists) and animals that drift with the currents. Together, these mostly microscopic organisms play a major role in maintaining the Earth system by, for example, carrying out almost half of the net primary production on our planet¹ and by exporting photosynthetically fixed carbon to the deep oceans^{2,4}. Plankton also form the base of food webs that sustain the complexity of life in the oceans and beyond⁵.

With the goal to gain a holistic understanding of this complexity, ocean ecosystems biology investigates how biotic and abiotic processes determine emergent properties of the ocean ecosystem as a whole⁶. Analogously to systems biology studies that require well-characterized cell lines or model organisms for a mechanistic, molecular understanding of their phenotypes, achieving this goal will require to establish an inventory of the ocean's plankton, to collect data on the interactions of organisms with each other and the environment, and to integrate this information in the context of physicochemical boundaries in the ocean ecosystem across space and time⁷. Global-scale efforts, although challenging, are poised to offer new insights into each of these directions

and should make possible better predictions of the impact of climate change on this crucial component of the biosphere.

Planetary-scale studies of open-ocean organisms have long been the stuff of dreams — from the *Challenger* Expedition (1872–1876), which led to the discovery and description of countless eukaryotic organisms, to the Global Ocean Sampling Expedition (2004–2008), which pioneered the genomic exploration of ocean microbial communities^{8–10}. Following this dream, *Tara Oceans* was conceived in 2008: a multidisciplinary project and team, including researchers with expertise in biological and physical oceanography, marine ecology, cell and systems biology, genomics, imaging as well as (bio)informatics, with a common goal to study epipelagic and mesopelagic plankton on a global scale (FIG. 1a) from the gene level to the community level⁷. At its beginning (BOX 1), this project, which would use the 36-m schooner *Tara* (FIG. 1b) for the expedition, required trade-offs and innovations in sampling needs and capabilities. Enormous planning was required to identify oceanic areas of scientific interest; to negotiate international waters, ports and sampling authorizations; and to resolve intense debates across disciplines to establish baseline sampling protocols. Finally, in September 2009, *Tara* set sail from Lorient, France, partially navigating through stormy weather and around pirates, to collect samples for analysis by

✉e-mail: ssunagawa@ethz.ch;
 vargas@sb-roscoff.fr
<https://doi.org/10.1038/s41579-020-0364-5>

state-of-the-art molecular and imaging technologies (Supplementary Box 1).

The primary objectives of *Tara* Oceans have been to generate a baseline understanding of plankton diversity, interactions, functions and phenotypic complexity across global taxonomic and spatial scales, and to communicate the scientific findings to the public and policymakers (BOX 2). In addition, all protocols, data and analyses (Supplementary Box 2) should be open access to promote further research. Working towards these goals, the consortium grew organically, and by the time the expedition was completed in 2013, *Tara* Oceans comprised 19 international partner institutions committed to generating, organizing and analysing the vast volumes of new and heterogeneous data derived from the thousands of plankton samples collected worldwide¹¹.

While complementary expeditions exploring the deep ocean as well as local-scale and other global-scale plankton surveys are under way^{12–16}, the focus of this Review is the work and outcomes of the *Tara* Oceans project. We describe how genetic, morphological and environmental data were combined and highlight the insights gained from the analysis of different plankton size spectra using a global ocean ecosystems biology approach. By providing an overview of the development of *Tara* Oceans from an adventurous initiative into a multinational, multidisciplinary, collaborative project, we also hope to stimulate planetary-scale research not only within a biome, but also across different ones, which will be crucial for integrating biology into models of Earth system functioning.

Author addresses

¹Department of Biology, Institute of Microbiology and Swiss Institute of Bioinformatics, ETH Zürich, Zürich, Switzerland.

²Department of Marine Biology and Oceanography, Institute of Marine Sciences-CSIC, Barcelona, Spain.

³Structural and Computational Biology, European Molecular Biology Laboratory, Heidelberg, Germany.

⁴Max Delbrück Center for Molecular Medicine, Berlin, Germany.

⁵Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany.

⁶Institut de Biologie de l'ENS, Département de Biologie, École Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France.

⁷Research Federation for the Study of Global Ocean Systems Ecology and Evolution, FR2022/Tara GOSEE, Paris, France.

⁸Université de Nantes, CNRS, UMR6004, LS2N, Nantes, France.

⁹Sorbonne Université, CNRS, Laboratoire d'Océanographie de Villefranche, Villefranche-sur-Mer, France.

¹⁰Stazione Zoologica Anton Dohrn, Naples, Italy.

¹¹Directors' Research, European Molecular Biology Laboratory, Heidelberg, Germany.

¹²Institute for Chemical Research, Kyoto University, Kyoto, Japan.

¹³PANGAEA, University of Bremen, Bremen, Germany.

¹⁴MARUM, Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany.

¹⁵Department of Microbiology, The Ohio State University, Columbus, OH, USA.

¹⁶Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, USA.

¹⁷Center for RNA Biology, The Ohio State University, Columbus, OH, USA.

¹⁸Génomique Métabolique, Genoscope, Institut de Biologie François Jacob, Commissariat à l'Énergie Atomique, CNRS, Université Evry, Université Paris-Saclay, Evry, France.

¹⁹Sorbonne Université and CNRS, UMR 7144 (AD2M), ECOMAP, Station Biologique de Roscoff, Roscoff, France.

*A list of authors and their affiliations appears at the end of the paper.

Sample collection and processing

From its departure in 2009 to its return in 2013, *Tara* sailed 140,000 km over a period of 38 months, systematically collecting more than 35,000 ocean water and plankton samples as well as environmental data (Supplementary Table 1) at 210 stations (FIG. 1). The aim was to sample most of the biogeographic and biogeochemical provinces¹⁷ of the global ocean and to follow standardized protocols and logistics for sample collection, distribution and storage to facilitate comparative analysis¹¹. To assess ocean plankton from viruses to small animals from the sunlit, epipelagic waters to the dark, mesopelagic waters down to 1,000 m, seawater was sampled with use of Niskin bottles, pumps and nets (FIGS 1c, 2A). Total plankton were then separated into fractions of organisms of different size ranges (FIG. 2B), and the fractions were cryopreserved on filter membranes or preserved in different fixatives for molecular and/or morphological analyses on land (Supplementary Box 3). Back in the laboratory, nucleic acids were extracted from the filters and subjected to high-throughput sequencing (HTS) to generate metabarcoding (metaB), metagenomic (metaG) and metatranscriptomic (metaT) data sets as well as to yield single-cell genomes (FIG. 2C). Deep sequencing was performed with Illumina technology at high coverage rates per sample to access the genomic content of plankton species, including those that are rare in the environment. As of 2019, more than 60 terabases from more than 2,800 individual plankton samples had been made publicly available (Supplementary Table 2). In addition, high-throughput imaging (HTI) tools captured the abundance and morphological features of plankton across size fractions spanning seven orders of magnitude (FIG. 2D). In total, 6.8 million images of planktonic organisms from more than 9,200 samples, amounting to more than 30 terabytes of data, have been generated (Supplementary Table 3).

The HTS and HTI data, combined with data from environmental conditions measured in situ and additional metadata (Supplementary Box 4), have been used for integrative analyses (see later and Supplementary Box 1) and released in open-source repositories (Supplementary Box 2), where they should serve as a treasure trove for future analyses. Furthermore, bioinformatic methods were either adopted or newly developed to facilitate the analysis, comparability and integration of the large volumes of data¹⁸. Still, despite these numerous bioinformatics resources and tools (Supplementary Boxes 5,6), powerful all-encompassing interfaces that make possible integration of the vast amounts of heterogeneous data generated by such global ecosystem-scale projects remain much needed for efficient secondary data usage.

Global ocean plankton biodiversity

Viruses. When *Tara* set sail in 2009, viruses were known to be abundant (one million to 100 million per millilitre of seawater), and were suggested to kill about one third of the microbial cells in seawater per day^{19–21}. These findings, however, were largely derived from counts of virus-like particles and incubation experiments. Over the last decade, a confluence of rapidly advancing sequencing technologies and low-input molecular

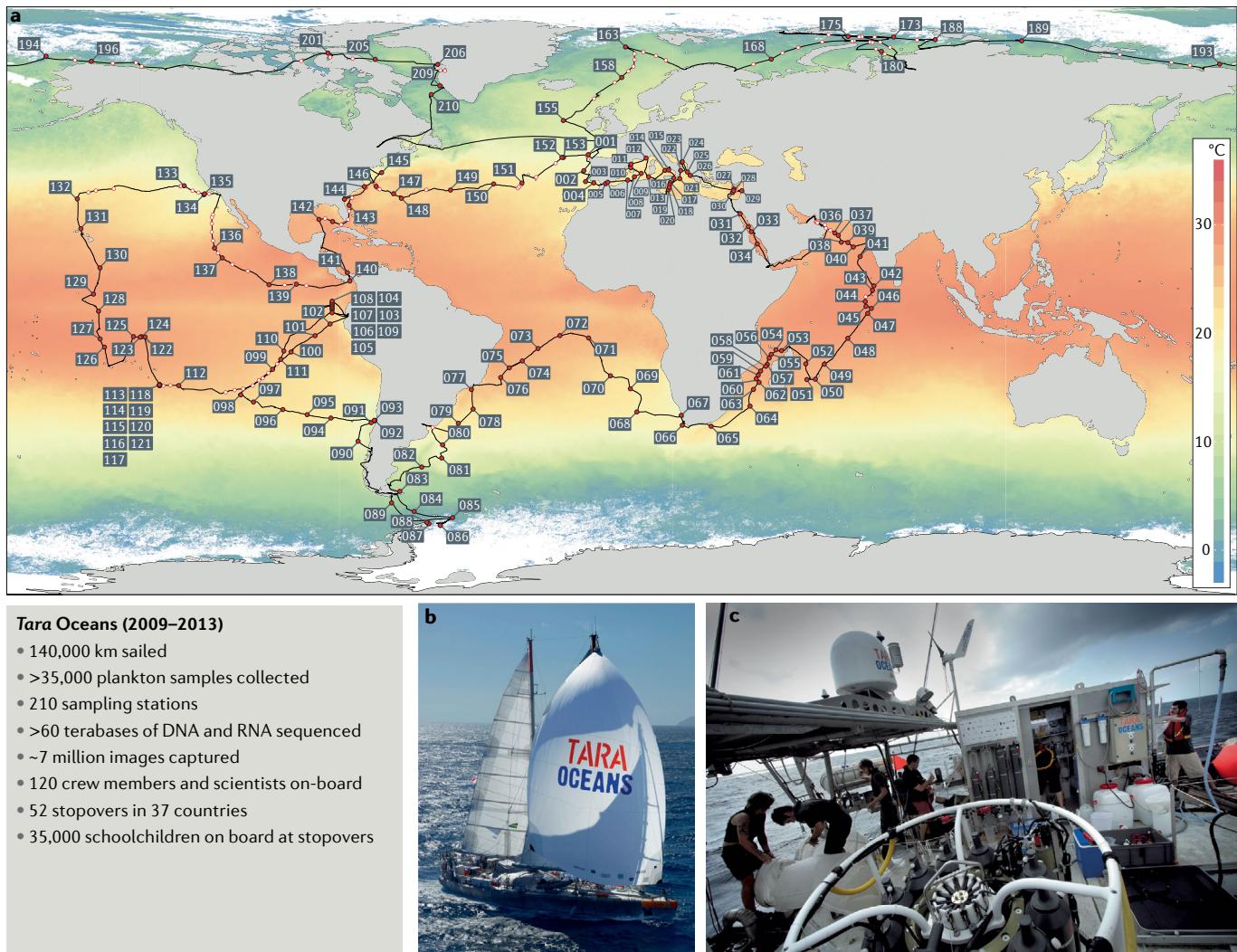


Fig. 1 | Tara Oceans sampled the global ocean ecosystem. **a** | The map shows the cruise track of *Tara* as she sailed the world from September 2009 to December 2013 and the location of 210 stations, which were chosen to cross and to sample as many biogeographic provinces and environmental features as possible (sea surface temperature shown as a colour gradient). Overall, more than 35,000 samples of seawater and plankton were collected and archived in partner laboratories. The samples are cross-referenced with the physicochemical data associated with each sample and sampling site (Supplementary Table 1). **b** | The 36-m aluminium-hulled schooner *Tara* hosted 15 crew members and scientists during the legs of her open-ocean sampling mission. **c** | In addition to the installation of a ‘dry laboratory’, mainly for imaging instruments, an on-board, ergonomic ‘wet-laboratory’ for plankton ecosystem sampling, from viruses to animals, was built on the rear deck of *Tara*.

biology techniques²² set the stage for systematic, quantitative global ocean virome surveys. These new capabilities advanced our knowledge of ocean viral genomes from 39 publicly available isolate genomes before *Tara* Oceans to metaG-derived sequence information for nearly 200,000 predominantly double-stranded DNA virus populations in the most recent global ocean virome (version 2; GOV2) data set^{23–25}.

This incrementally increasing (FIG. 3a) data set set the stage for new approaches to the taxonomy of double-stranded DNA viruses that infect bacteria and archaea. Population genomic analyses and ecological phenotyping suggest, at least for culturable phage–host model systems, that these hundreds of thousands of virus populations across the global ocean represent

a taxonomic rank of ‘species’. For one, this conclusion is based on the notion that gene flow is higher within than between the populations of cyanophages that were deeply sampled from coastal Pacific Ocean seawater and from evolutionary selection analyses that suggest that most of the populations are under differential selection pressures²⁴. Among ‘heterotrophic phages’, however, the populations had measurably differing niches as inferred from host-range differences²⁶. Furthermore, at least for viral populations that were assembled from short-read metaG data, the population-delineating benchmark of 95% average nucleotide identity differentiated more than 99% of the virus populations in the GOV2 data set²⁴. Undoubtedly, there are microdiverse populations that cannot be assembled completely from such data sets,

Heterotrophic
Capable of incorporating organic carbon into biomass.

Remineralized

Derived from the breakdown of organic matter into its simplest inorganic form.

as inferred from emerging single-virus genomics²⁷ and long-read viromics measurements²⁸, but the prevalence of such populations remains unclear. Beyond the species level, the scale of the data necessitated an automated and systematic approach to organize the viral sequence space at the level of viral genera^{29,30}, resulting in the modernization of the gene-sharing network-based taxonomy^{31,32} that is currently a leading tool for classifying the ‘dark matter’ that dominates the virosphere^{33,34}.

Once taxonomically organized, the data provided a first glimpse into large-scale ecological patterns and drivers for ocean viruses. For example, viral communities seem structured (likely indirectly through their hosts) by temperature and oxygen, and are passively transported by oceanic currents, consistent with the notion that ‘everything is everywhere and the environment selects’^{23,35}. Viral ecology patterns were also revealed at the between-population (macrodiversity)

and within-population (microdiversity) levels, the latter tracking more recent ecological and evolutionary changes. These data, derived from the GOV2 data set²⁴, revealed not only that the oceans globally comprise five ecological zones but also latitudinal biodiversity patterns that are both consistent with (low in the Southern Ocean, at least at the northernmost margins sampled by *Tara* Oceans, and high at the equator) and contrast with (surprisingly high in the Arctic) those known for macroorganisms^{24,36} (FIG. 3b). Beyond these large-scale patterns, ocean viruses have now also been linked in silico to ecologically important marine microbial hosts, providing foundational hypotheses that can be tested by experimental virus–host linkage methods. These data have provided global virus–host infection maps²⁵, and for cyanobacteria they have advanced our understanding of how viral infections associate with diel dynamics of host communities^{37–39}. Specifically, a cross-omics analysis that leverages *Tara* Oceans viral genome data revealed a peak of cyanophage gene expression in the afternoon or dusk followed by an increase of genomes from the virions at night, confirming that cyanophages drive the diel release of cyanobacteria-derived organic matter into the environment³⁹. The data have also revealed key virus-encoded auxiliary metabolic genes that indicate extensive metabolic reprogramming of the hosts and are likely to directly modulate biogeochemical cycles^{25,29,40}. These genes range from photosynthesis genes, which were identified in cyanophage isolate genomes more than a decade ago^{41,42}, to genes that manipulate central carbon, sulfur and nitrogen metabolism^{25,43}.

From an ecosystem perspective, these findings question some paradigms in phage and ocean microbiology. For example, a multi-omics mechanistic study of a cultured representative of the third-most-abundant ocean viral genus (*Bacteroidetes* phage phi38:1) revealed that these viruses infect two bacterial strains with identical 16S ribosomal RNA gene sequences in completely different ways due to a diversity of mismatched metabolic machinery and ameliorated cellular defences. These data suggest that phage resistance in nature is not due to simple, single-step mutational events but is rather due to a multistep and more complex process⁴⁴. A solid understanding of the ecosystem outputs of cells infected by viruses (‘virocells’) remains elusive. To address this knowledge gap, an experimental model system using viruses that infect *Pseudoalteromonas*, the second-most highly predictive bacterial genus for ocean carbon flux⁴⁵, was investigated by a multi-omics mechanistic approach, which revealed that virocells differ drastically from uninfected cells and that virocells infected with one phage are completely different from those infected with another phage⁴⁶.

Additionally, the vast *Tara* Oceans organismal data set coupled with global measurements of ocean carbon export provided an opportunity to determine which organisms best predicted this crucial ecosystem function. For decades, the paradigm in viral ecology has been that viruses ‘keep carbon small’, as lysis products are rapidly remineralized⁴⁷; however, early observations that viruses seemed to sink, as inferred from photosynthesis gene sequences at various depths⁴⁸ and later

Box 1 | A historical account of the *Tara* Oceans expedition

Tara Oceans was conceived by Eric Karsenti to popularize fundamental science using a sailing boat. The *Tara* Ocean Foundation proposed the use of its schooner *Tara* for a global expedition. A scientific component focused on plankton was soon added through inputs from Christian Sardet and Gaby Gorsky in 2007 (zooplankton) and Chris Bowler and Colomban de Vargas in 2008 (microbial plankton). The idea was shared with other scientists, leading to the development of an international, multidisciplinary, collaborative consortium aimed at studying oceanic plankton at a planetary scale.

Development from a rough concept to a project of its current magnitude required a coalescence of many factors. Through 2008, a team of *Tara* Oceans coordinators with complementary expertise began to grow. New members joined the project through word of mouth, and regular meetings were held, approximately every 3 months, to define the structure of the project. This crucial start-up phase was made possible through seed funding from the French National Research Agency (ANR), the French National Centre for Scientific Research (CNRS), the European Molecular Biology Laboratory (EMBL), the Veolia Foundation and Region Bretagne, which recognized the potential of the project. During this time, the overall collaborative philosophy of the project, the holistic and systematic sampling strategy and the details of the sampling and analysis protocols were established. Meetings for project coordination and networking continued over the last 11 years, rotating between Paris, Roscoff (France) and Heidelberg (Germany), among other locations.

The principles for the consortium were modelled on the basis of a scientific unit at EMBL, in which group leaders from different disciplines with interests in the same broad scientific question meet regularly and structure projects with a bottom-up approach. Similarly, for *Tara* Oceans, decisions were often made on the fly on the basis of discussions between the coordinators and were overseen by a programme manager, Stefanie Kandels. This form of planning represents an entirely different and more agile type of science than what is generally supported by peer-reviewed funding bodies that require an *a priori* statement of the research design and goals. Although riskier, this blue-sky approach offers opportunities to develop creative ideas that may lead to novel and innovative research directions.

Furthermore, *Tara* Oceans is an example of how adventurous science can profit from engaged philanthropists and private entities, such as agnès b., the *Tara* Ocean Foundation and other private foundations and companies (see Acknowledgements), to catalyse a new approach for supporting fundamental biological research. Importantly, *Tara* Oceans consortium members invested their own resources in the project, which did not necessarily fit into public mainstream channels of science funding. Finally, as the project gained momentum and credibility, funding from national agencies, including the French Government through its Investissement d’Avenir programme (project OCEANOMICS), was acquired, covering the substantial costs associated with the processing and sequencing of all samples as well as the general running costs of the project. Major catalysts of the project were the flexibility and cost-effectiveness provided by use of a sailing boat to collect plankton samples across the global ocean, and the subsequent use of the most advanced technologies in sequencing, imaging, data analysis and computing onshore at the multiple partner institutions involved.

Box 2 | Outreach and societal impact

The mission of the Tara Ocean Foundation includes outreach that makes possible the combination of innovative science with diverse activities to communicate the project goals and findings to the public, company managers, policymakers and schoolchildren. During the expeditions from 2009 to 2013, *Tara* completed 52 stopovers in 37 countries. In each of the ports of call and in the city of Paris, where *Tara* was docked during the 2015 United Nations Climate Change Conference, the crew and scientists welcomed several thousand visitors onboard. More than 50 exhibits explained the goals of the project and demonstrated the key role that tiny organisms drifting in the oceans play in the global ecology of our planet. Tours aboard *Tara* and the numerous conferences presented by members of the team in every country were inspirational to the visitors, who included local decision makers (mayors, ministers, heads of states and the United Nations Secretary General) in addition to many schoolchildren. The story of scientists, sailors and inspired artists criss-crossing the planet on a schooner to explore the ocean using state-of-the-art technologies also attracted non-scientists and the media around the world. Photographs and videos received worldwide attention (for example, *plankton chronicles* and *artist profiles*, and journalists wrote articles about the expedition from many different angles, especially following the publication of major scientific results in 2015 (REF.¹⁴⁹). In addition, the *Tara Ocean Foundation* published three journals in French, English, Chinese, Japanese and Portuguese, and the expedition was popularized through documentaries on television (for example, 35 prime-time *Thalassa* shows on French television in 2009 and 2010) as well as several DVDs and books.

gene-to-ecosystem modelling predictions⁴⁵, suggested bacterial and archaeal viruses are key players in the biological carbon pump, at least in the predominantly open-ocean waters that were sampled. Specifically, the abundances of viruses best predicted global ocean carbon flux in comparison with abundances of bacteria and archaea or eukaryotes, and a handful of the most predictive viruses were identified to guide future work⁴⁵. Although improved modelling techniques are needed to better capture the magnitudes of variation in carbon flux and to simultaneously compare the relative impact of all organismal groups, these gene-to-ecosystem modelling predictions suggest that viruses are important to an ecosystem process that was traditionally viewed to be dominated by the mere presence of large protists and metazoans.

Furthermore, *Tara Oceans* deep-sequencing data provide unprecedented opportunities to unveil the global eukaryotic virosphere^{49–52}. Although eukaryotic viruses are less abundant than bacterial and archaeal viruses, gene marker-based approaches have revealed that eukaryotic viruses are at least as abundant as archaea in the epipelagic layer of the ocean⁵⁰. Both metaG data and metaT data suggested that nucleocytoplasmic giant DNA viruses are ubiquitous and transcriptionally active across oceans¹⁹. Reminiscent of the phage-to-host ratio, these giant DNA viruses outnumber the abundance of their potential hosts by an order of magnitude⁵⁰ and show dispersal at a planetary scale⁵³. Moreover, the taxonomic richness of these giant viruses was shown to be potentially greater than that of bacteria and archaea⁵⁴.

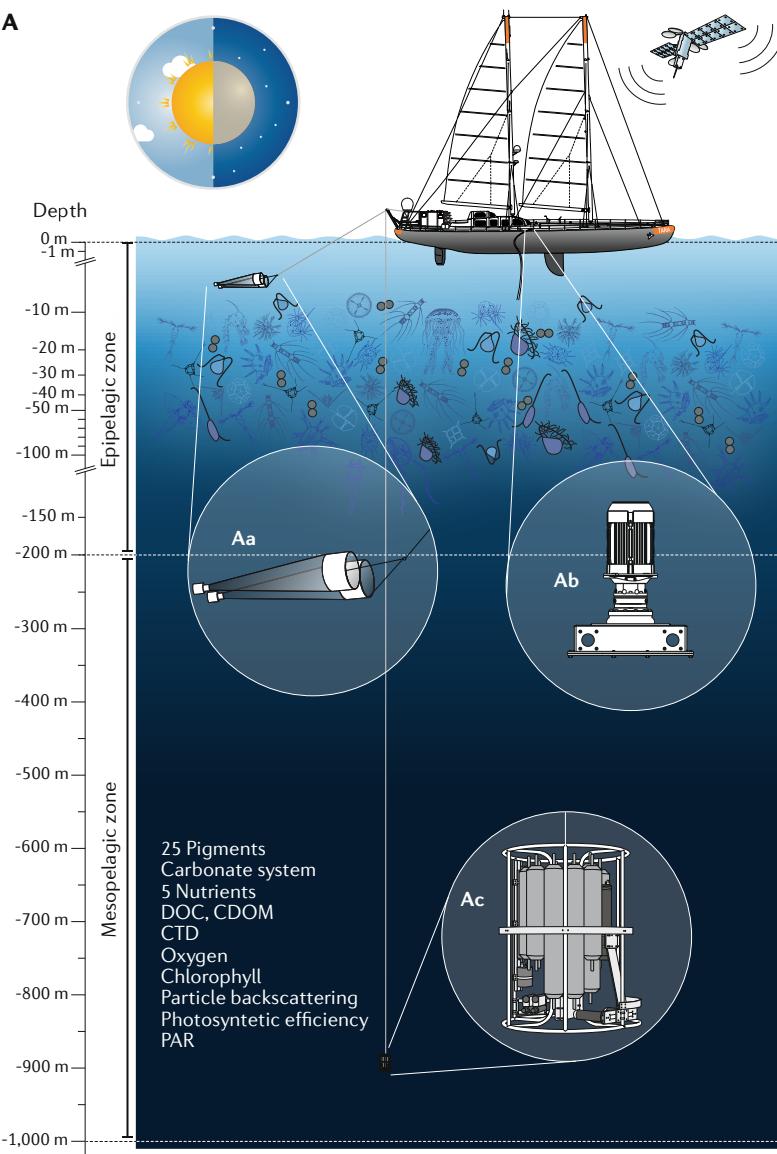
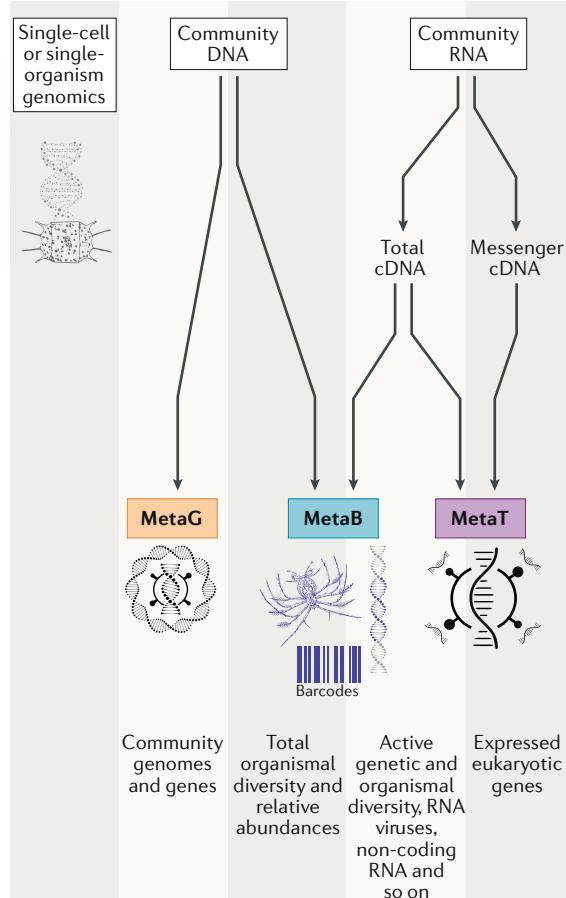
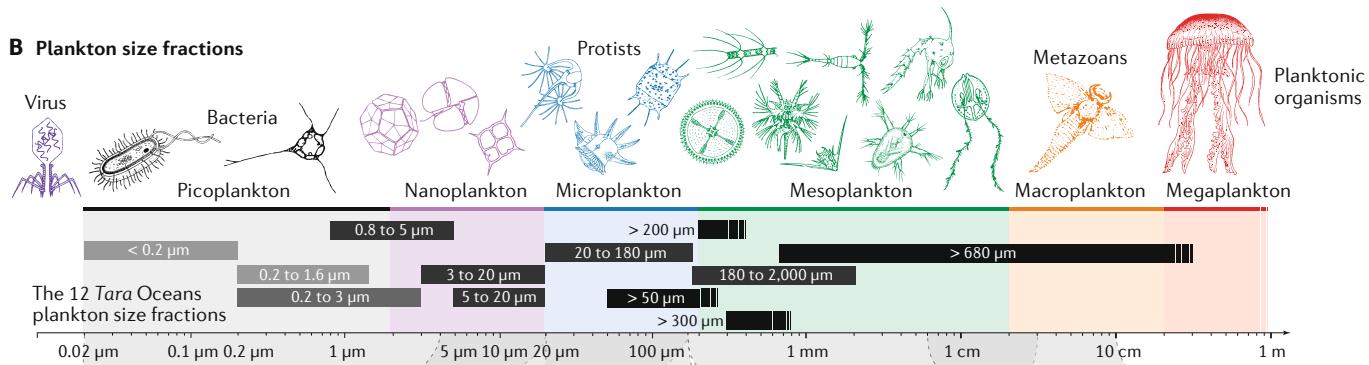
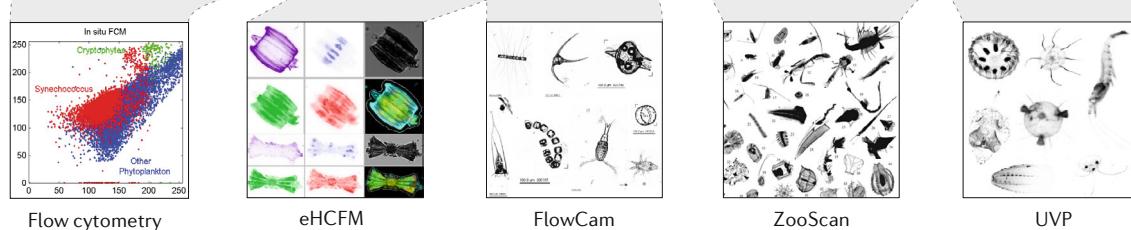
The study of the associated genomes has also led to an improved understanding of the impact of eukaryotic viruses on the evolution of the host's sexual life cycle. For example, the loss of the genomic capacity to carry out a sexual life cycle in the cosmopolitan phytoplanktonic organism *Emiliania huxleyi* in the oligotrophic ocean was found to be associated with decreased biotic pressure due to the low abundance of large virus infection⁵⁵. Environmental factors, such as phosphate availability,

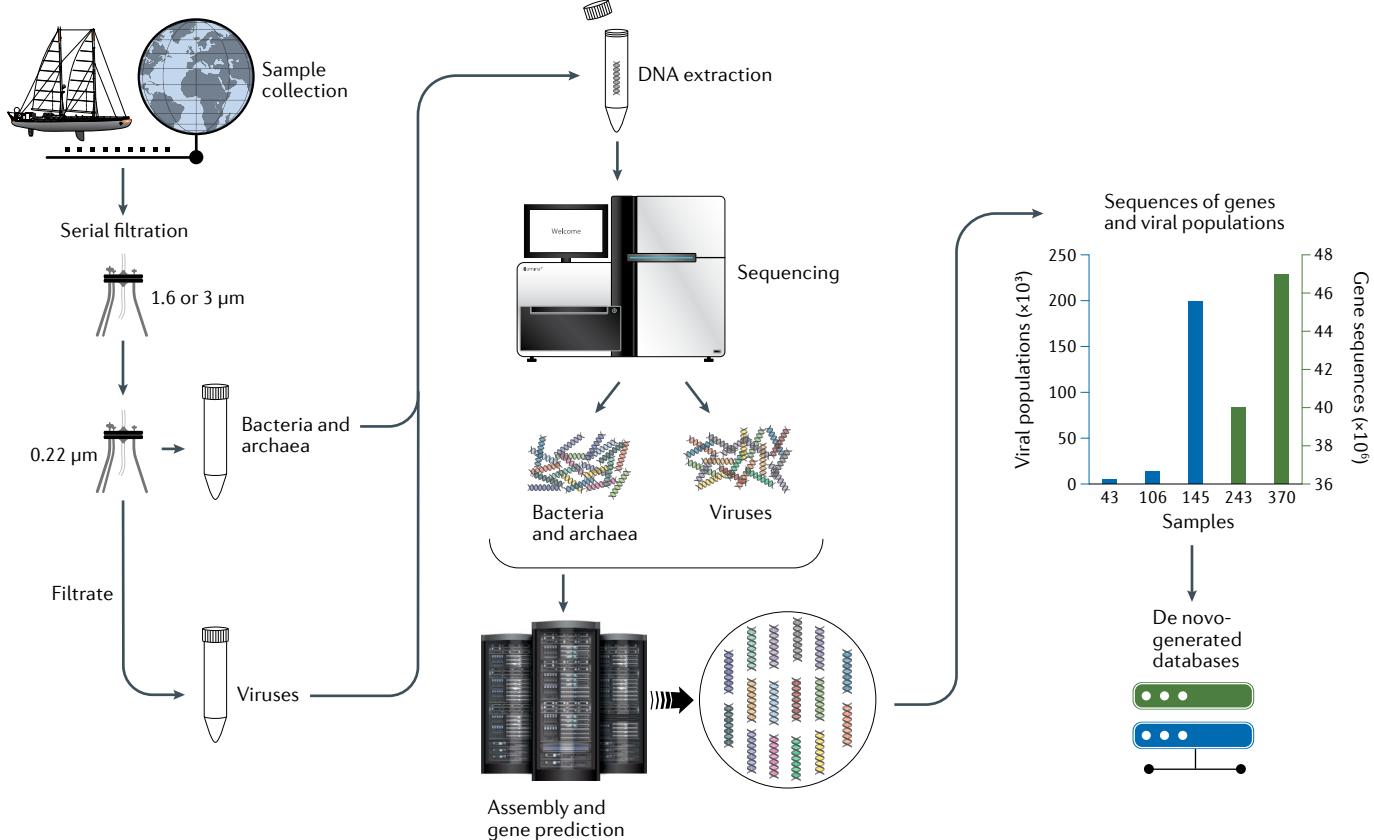
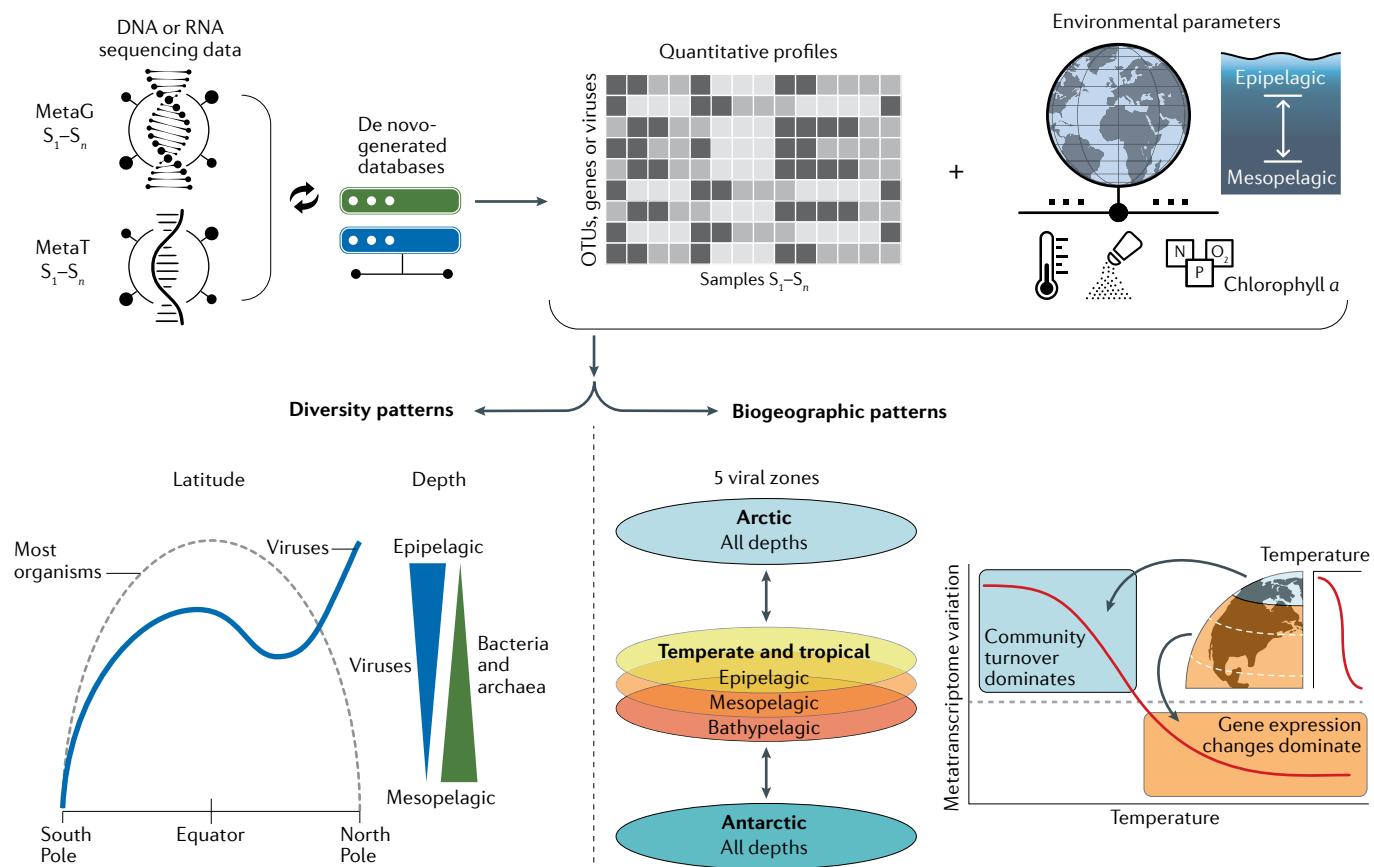
were further found to drive giant virus community structure in some regions of the ocean^{56–59}. Furthermore, some eukaryotic viruses are predicted to influence the efficiency of the vertical carbon flux⁶⁰.

Overall, these studies led to the development of community resources, including iVirus, which provides access to viromic tools and data sets⁶¹ and a knowledge base of virus–host interactions⁶², to facilitate further eco-genomic exploration of marine viruses. Of note, single-stranded DNA viruses and RNA viruses from the *Tara Oceans* samples and data sets have yet to be analysed. Although single-stranded DNA viruses are not thought to be abundant in marine systems²⁵, RNA viruses are suggested to constitute as much as half of the viral particles in the oceans⁶³.

Bacteria and archaea. The Global Ocean Sampling Expedition pioneered the exploration of the genomic diversity of ocean bacteria and archaea on the basis of environmental DNA sequencing^{8,10}. Analysis of surface water samples collected at 41 locations from the eastern

Fig. 2 | *Tara Oceans* assessed plankton across taxonomic, organismal and environmental scales to study the whole ecosystem. **A** | At each open-ocean station (FIG. 1), *Tara* sampled plankton during daytime and night-time, guided by satellite data, using five types of plankton nets with different mesh sizes (part **Aa**), an industrial, high-volume peristaltic pump (part **Ab**) and a rosette water sampling system equipped with Niskin bottles (part **Ac**), from sunlit (surface and subsurface, including the deep chlorophyll maximum) to dark (mesopelagic) waters down to 1,000-m depth. Key physicochemical parameters of the sampled water were measured in situ or back in the laboratory (Supplementary Table 1). **B** | The *Tara Oceans* sampling protocol targeted in total, although not at every sampling site, 12 organismal size fractions from picoplankton to megaplankton, that is, across more than seven orders of magnitude in size, corresponding to the range from the size of a bee to ten times the height of Mount Everest. **C** | The *Tara Oceans* high-throughput sequencing workflow generated multi-omics data sets (Supplementary Box 1; Supplementary Table 2) for assessment of the diversity and relative abundance of genomes, genes and taxonomic barcodes across the kingdoms of life. **D** | The high-throughput imaging methods applied by *Tara Oceans* imaged plankton from different size fractions (Supplementary Table 3) to quantify organismal richness, sizes, biovolumes and morphological complexities. Owing to field work conditions and prioritization of specific analyses, it was not possible to collect every sample at each station and to subject every sample to all possible types of analyses. However, each sample is cross-referenced to a rich set of metadata to provide researchers with the possibility to ensure comparability of different samples and data types. All physicochemical, sequencing and imaging data obtained are archived following FAIR (findable, accessible, interoperable and reusable) principles to facilitate integrative analyses (FIG. 6). CDOM, coloured dissolved organic matter; CTD, conductivity, temperature and depth; DOC, dissolved organic carbon; eHCFM, environmental high-content fluorescence microscopy; metaB, metabarcoding; metaG, metagenomics; metaT, metatranscriptomics; PAR, photosynthetically active radiation; UVp, underwater vision profiler. Parts **B** and **D** adapted from REF.⁷, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>).

A**C High-throughput sequencing****B Plankton size fractions****D High-throughput imaging**

a Resource generation**b Data analysis**

◀ Fig. 3 | **Tara Oceans viral, bacterial and archaeal analysis pipeline and highlighted discoveries.** The analysis of sequencing data from environmental DNA enriched for viruses, bacteria and archaea and environmental RNA enriched only for the last two involved the establishment of reference resources that were subsequently used to study diversity and biogeographic patterns. **a** | Reference resources (for example, seawater samples) were sequentially filtered to separate plankton into several size fractions. For bacteria- and archaea-enriched samples, seawater was filtered through 1.6- or 3- μm filters and collected on 0.22- μm filters. Viruses were flocculated in 0.22- μm filtrates using ferric chloride and collected on 1- μm filters¹¹. On DNA extraction, library preparation and sequencing, DNA sequencing reads were assembled into contigs. For viruses, contigs were screened for sequences of viral origin and then grouped for individual viral populations^{24,25,149}. For bacteria and archaea, genes were predicted on contigs and clustered to yield catalogues of non-redundant gene sequences^{64,65}. **b** | These de novo-generated resources were used as reference databases for quantifications of viruses, genes and microbial species per sample (S). Bacterial and archaeal quantifications were derived as abundances of operational taxonomic units (OTUs) based on 16S ribosomal RNA fragments directly identified from metagenomic (metaG) sequencing reads¹⁵⁰. Integration of quantitative profiles with environmental parameters (Supplementary Table 1) facilitated the study of diversity gradients across latitude and depth, with partly contrasting patterns observed for viruses compared with bacteria and archaea^{24,64}. Biogeographic analyses revealed five ecological zones of viral populations²⁴, and differences in the mechanisms driving community transcriptomic compositions were identified by combining metaG and metatranscriptomic (metaT) data⁶⁴. Part **b**, bottom right adapted from REF.⁶⁴ and bottom left from REF.²⁴, Elsevier.

North American coast through the Gulf of Mexico and into the equatorial Pacific revealed approximately six million new protein sequences, almost doubling the number available in public databases in 2005 (REF.¹⁰). Despite this prompt expansion of ocean microbial protein sequences, the diversity of protein-coding genes in nature was too large to be captured with the sequencing technology available at the time. Specifically, new protein families were discovered at nearly linear rates with additional sequencing¹⁰.

Because of the advent of HTS technologies and the drastic reduction of costs since 2008, *Tara Oceans* generated unprecedented amounts of environmental sequencing data for each sample with the goal to obtain an ecosystem-wide overview of the diversity, function, biogeography and activity of the global ocean microbiome^{64,65}. For a set of 243 samples enriched in planktonic organisms smaller than 3 μm , more than 7.2 terabases of metaG sequencing data were assembled into an ocean microbial reference gene catalogue (FIG. 3a) by use of a method originally developed for human microbiome research^{66–68}. This first ocean gene catalogue comprised more than 40 million non-redundant protein-coding sequences. This number was four times higher than the one reported for the human gut microbiome at the time⁶⁵, although approximately two thirds of the gene abundances could be attributed to core gene families that were shared between the two biomes. Moreover, 80% of these sequences were previously unknown, on the basis of a nucleotide sequence identity of more than 95% with sequences in reference databases. The rate of detecting new genes from an estimated 35,000 bacterial and archaeal operational taxonomic units (OTUs; 97% clusters) decreased to 0.01% by the end of sampling, suggesting near saturation of gene diversity in these samples.

These newly established resources have thus substantially expanded our knowledge of the ocean microbial gene repertoire and made possible taxonomic and gene functional composition analyses of ocean microbial

communities on a global scale⁶⁵. In agreement with prior studies^{69,70}, microbial communities sampled worldwide at 68 locations from epipelagic and mesopelagic waters were primarily structured by depth. In addition, the taxonomic and gene functional diversity was higher in mesopelagic layers than in epipelagic layers, whereas viruses showed the opposite pattern for the same latitudinal range²⁴ (FIG. 3b). Beyond depth, previous studies suggested temperature and other factors, such as salinity and nutrients, as important drivers of the taxonomic composition of ocean microbial communities⁷¹. Because of global sampling by *Tara Oceans*, it was possible to disentangle geographic effects from environmental effects (that is, the similarity of microbial community compositions may be driven by geographic proximity rather than environmental similarity of the respective sampling locations), and as a result pinpoint temperature as a key variable to predict the taxonomic and gene functional composition in epipelagic waters of the open ocean⁶⁵.

In addition, the availability of metaT data from 126 sampling sites, including the Arctic Ocean (FIG. 1a), made it possible to address the question of how microbial communities adjust to global environmental variation. Such adjustments seem to differ not only for individual metabolic processes but also for oceanic regions⁶⁴. These conclusions were reached through the integration of metaT and metaG data, along with the development of new bioinformatics resources and normalization procedures. Specifically, a new ocean microbial gene catalogue with 47 million sequence entries was generated to facilitate the integration of metaT and metaG data for gene-level quantitative analyses of community transcript, gene abundance and gene expression levels. Normalization procedures based on single-copy marker genes^{64,72} made it possible to distinguish organismal turnover and gene expression changes as underlying mechanisms for changes in the pool of community transcripts. These analyses suggested that polar communities are more specifically adapted to their niches and may undergo stronger organismal composition changes than their physiologically more variable counterparts in temperate and tropical waters in response to rising seawater temperatures⁶⁴ (FIG. 3b).

More focused analyses of *Prochlorococcus* and *Synechococcus*, the two most abundant and widespread phototrophic bacterial genera in the ocean, have led to a better understanding of their genetic capacity for mixotrophy as well as factors that control their biogeographic distribution^{73–75}. In addition, single-cell genome sequencing revealed a novel species in the genus *Kordia* (phylum Bacteroidetes) with the potential capacity for photoheterotrophy^{76,77}, and combined HTS and morphological analyses contributed to identifying the functional and ecological importance of a symbiosis between nitrogen-fixing cyanobacteria and haptophytes in epipelagic waters worldwide^{78–80}.

Finally, the resources and data provided by *Tara Oceans* have also made possible research by investigators outside the consortium. For example, following the first release of metaG data⁶⁵, thousands of draft genomes, so-called metagenome-assembled genomes (MAGs), from ocean bacteria and archaea were reconstructed^{81–84}.

Mixotrophy
Capacity to incorporate carbon into biomass from either inorganic or organic sources.

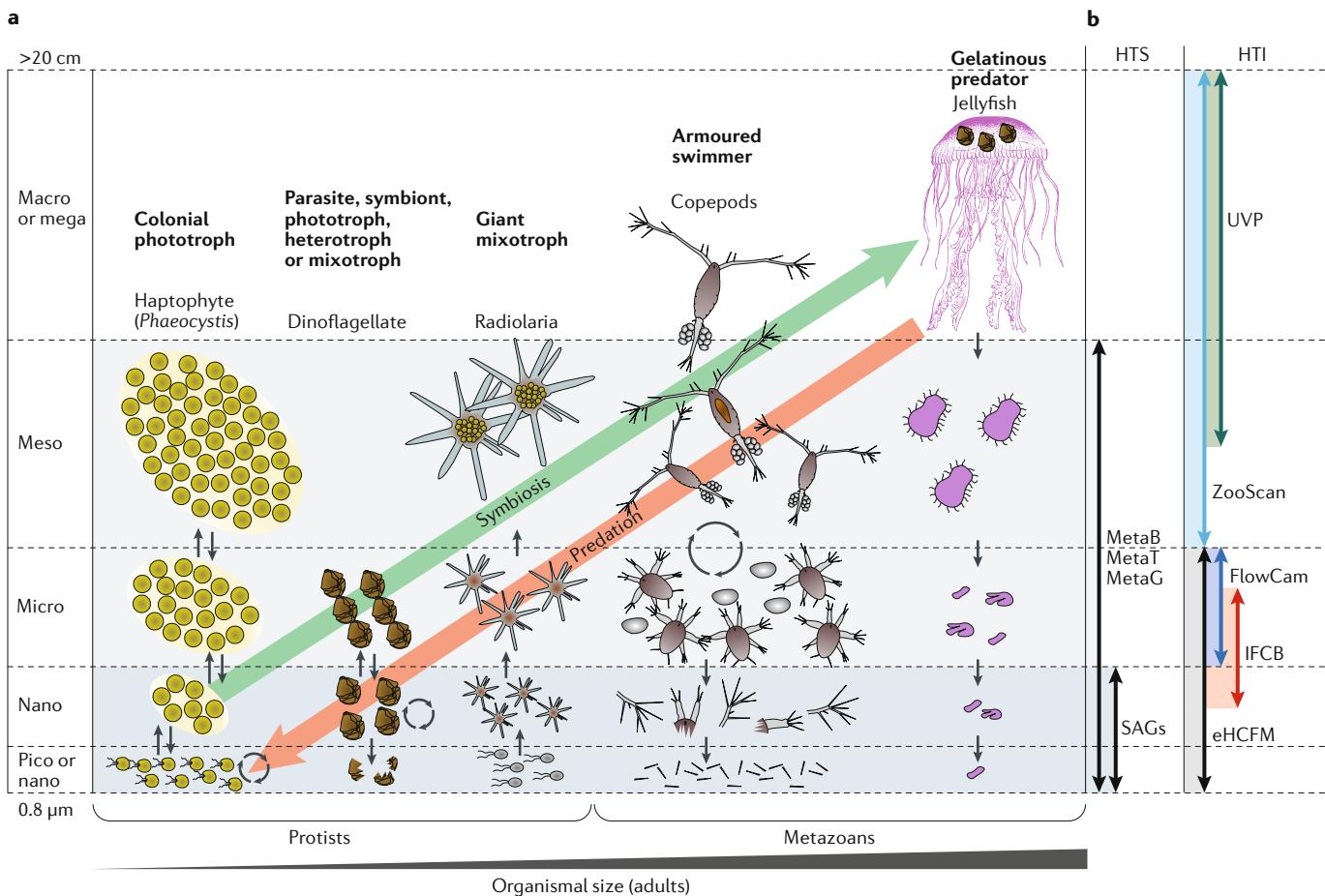
Photoheterotrophy
Capacity to derive energy from light and carbon from organic matter.

Haptophytes
Group of single-celled photosynthetic planktonic organisms.

Metagenome-assembled genomes (MAGs). Consensus genome sequences that are reconstructed using sequencing reads of DNA extracted from whole microbial communities.

Some MAGs shed new light on the metabolic capabilities of certain bacterial lineages³¹. Specifically, a number of bacteria in the phylum Planctomycetes, a group that had previously not been linked to nitrogen fixation, were found to contain and express³⁴ genes that are required for this process. Other MAGs led to an extended view of the phylogenetic diversity of Alphaproteobacteria, questioning the long-standing hypothesis about the origin of mitochondria. Rather than mitochondria having evolved

from the alphaproteobacterial order Rickettsiales, the authors suggested that mitochondria may have evolved from a lineage that branched off before the divergence of all Alphaproteobacteria sampled to date³². Other studies discovered new families of light-sensing rhodopsins³⁵, including a new class of anion-conducting channel rhodopsin³⁶. Furthermore, Tara Oceans data were used in other ways, such as to define metabolic functional groups and to decouple their global variation from taxonomic



Eocene epoch

Second geological epoch of the Palaeogene period (66 million to 23 million years ago) that began 56 million years ago and ended 34 million years ago.

Southern Ocean gateways

Pathways of the oceanic circulation that are influenced by the displacement of continents (for example, the Drake Passage, South African gateway and the Tasman gateway between Antarctica and South America, Africa and Australia, respectively).

community composition⁸⁷, and to estimate the contribution of marine organisms to the total biomass on Earth⁸⁸. These examples showcase the accomplishment of one of the initial goals of the project: the wide use of *Tara Oceans* data by the broad scientific community to make possible new discoveries and follow up on diverse research questions.

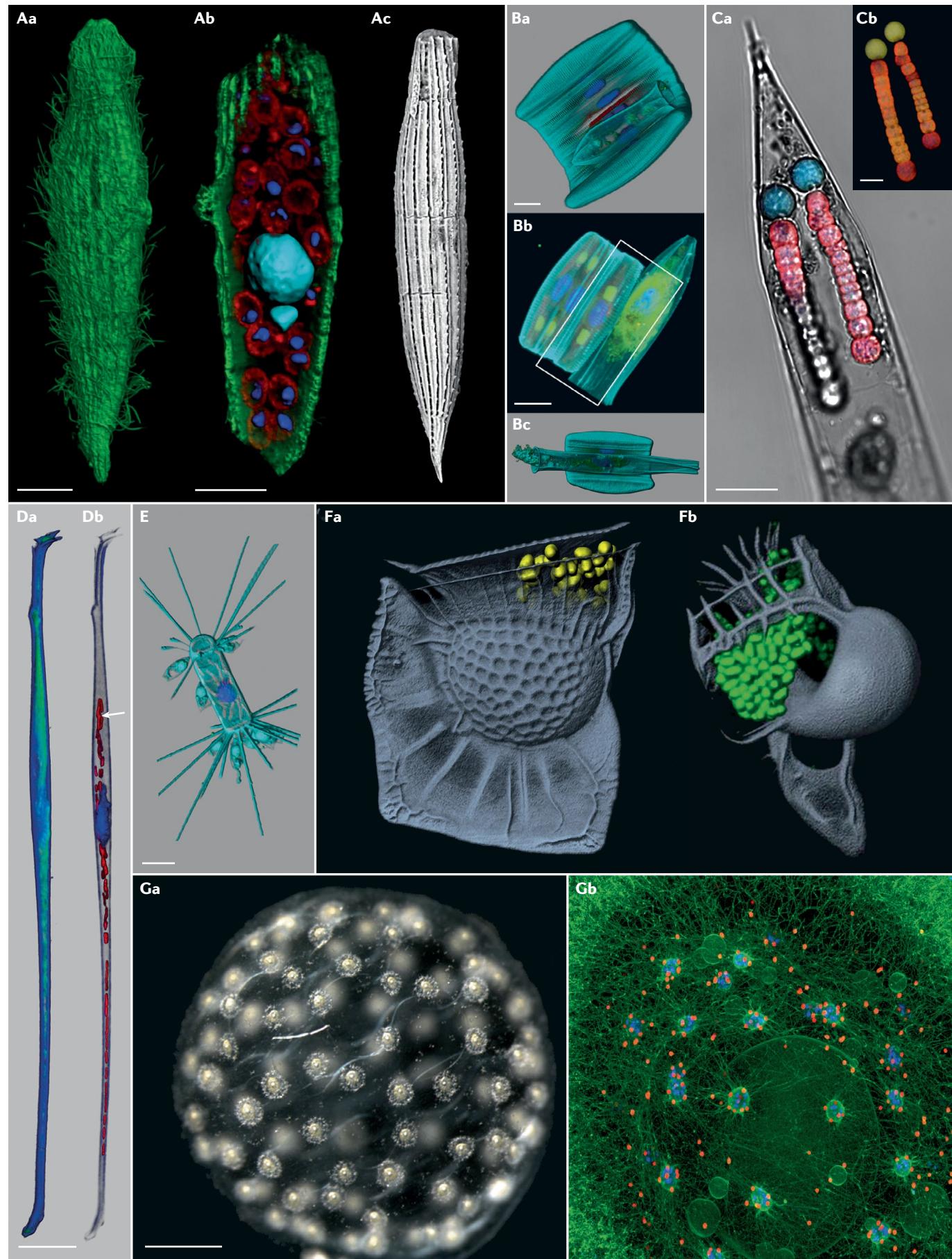
Protists. Protists, which span a wide range of organismal sizes from less than 1 µm to a few millimetres, can have extremely large genomes and complex morphologies, physiologies and behaviours⁸⁹ (FIG. 4). In 2009, knowledge of planktonic protists was based mainly on DNA metabar and flow-cytometry surveys for the smallest taxa, microscopic observations for the larger ones (larger than 20 µm) and genomic characterization of a handful of model organisms, largely phototrophic species (phytoplankton), including alveolates, diatoms, coccolithophores and prasinophytes. In oceanography, protists are traditionally divided on the basis of their broad ecological function into phytoplankton, heterotrophic nanoflagellates and larger predators. To expand this incomplete knowledge base, *Tara Oceans* developed an automated high-resolution 3D imaging workflow for quantitative subcellular exploration of microeukaryotes⁹⁰ and generated more than 220 billion DNA sequencing reads from about 2,200 samples (Supplementary Table 1). Furthermore, 6.8 million images from ~9,000 eukaryote samples covering organismal size fractions from 0.8 µm to a few centimetres (picoplankton and nanoplankton to macroplankton and megaplankton; Supplementary Table 2) were obtained (FIGS 4b,d,5). This data set of protist biocomplexity was completed with the building of new reference resources of taxonomically curated DNA barcodes^{91,92}, transcriptomes⁴⁹ and single-cell genomes^{93,94}.

The first large-scale metaB survey based on a fragment of the 18S ribosomal RNA gene⁹⁵ revealed about 150,000 eukaryotic taxa (genus or higher taxonomic levels) in the epipelagic ocean, and only ~10% of these taxa were known previously. More than 85% of these taxa represent uncharacterized protists of mostly heterotrophic groups⁹⁶, including many parasites and symbionts^{78,79,97–99} (FIG. 4c), in addition to the traditional members of the plankton community (such as diatoms¹⁰⁰, dinoflagellates¹⁰¹, prasinophytes¹⁰² and ciliates¹⁰³).

The *Tara Oceans* metaB survey has become a reference baseline for the community to assess global upper ocean diversity and biogeography of specific taxa or functional groups^{104–106} and as a test data set for new bioinformatic tools^{107,108}. Targeted analysis of the major eukaryotic phytoplankton groups (diatoms, dinoflagellates, haptophytes, pelagophytes and chlorophytes) has clarified their relative abundances with respect to each other as well as with respect to mixotrophs and known photosymbionts and with respect to the different organismal size fractions collected by *Tara Oceans*¹⁰⁹. Clade-specific analyses across plankton size fractions further revealed the importance of nano-sized and pico-sized diatoms that had previously been overlooked in ocean surveys¹¹⁰. These minute diatom species, including *Minidiscus* spp. and *Minutocellus* spp., were found to be globally distributed, and data from the DeWeX cruise in the Mediterranean Sea revealed that these organisms can generate massive blooms and can also be found at depth, implying a substantial contribution to carbon export¹¹¹. In addition, the combination of metaB data with palaeoenvironmental data and phylogenetic models of diversification were used to analyse the evolutionary diversification of the entire group of diatoms. There was a negative correlation between carbon dioxide partial pressure and early diatom diversification, consistent with increased primary productivity (that is, conversion of inorganic carbon into organic carbon) that favours increased diversity. Subsequently, in the late Eocene epoch, a major burst of diversification occurred at around the same time as the Southern Ocean gateways opened, creating a new ecosystem where diatoms could thrive. The molecular data are consistent with previous reports based on analysis of diatom microfossils¹¹². This diversification was affected by changes in sea level, an influx of silica and competition with other planktonic groups, and different diatom clades were affected differently. This heterogeneity in diversification dynamics across diatoms suggests that a changing climate will favour some clades at the expense of others⁹⁴.

Furthermore, the deep coverage of the *Tara Oceans* metaB data sets (typically one million to two million sequence reads per sample) has made possible exploration of the rare protist biosphere⁹². Briefly, an adaptive algorithm was used to explore the variant abundance distributions of non-dominant OTUs across plankton communities. These rare OTUs constituted more than 99% of the local richness in each sample, and their relative abundances were governed by a power law. Despite the apparently very high spatial turnover in species composition at a given site in the ocean, the power-law exponent varied by less than 10% across locations and showed no biogeographic signature. Such

◀ Fig. 4 | *Tara Oceans* analysis of eukaryotic plankton complexity and highlighted discoveries. **a** | This illustration shows the biological and functional complexity of eukaryotes across the plankton organismal size fractions analysed in *Tara Oceans*. Whereas tiny phytoplanktonic organisms (for example, *Phaeocystis*) can assemble into visible colonies, heterotrophic protists in association with phytoplanktonic organisms (for example, *Collodaria*) can form giant holobionts, which outweigh all animals in (sub) tropical sunlit oceans¹¹⁷. On the other hand, animals produce gametes, juvenile stages and debris which might be important components of microbial plankton size fractions. Overall, eukaryotes show diverse and complex interactions and behaviours along the symbiosis and predation axes, reflecting their immense and non-saturating gene repertoire¹⁹. Note that the viral, bacterial and archaeal diversity associated with eukaryotes is not represented in this scheme. **b** | Different molecular and imaging methods were developed and/or deployed by *Tara Oceans* to explore and assess unicellular and multicellular eukaryotes across their ontogenetic, life-cycle and symbiotic complexity. **c** | This schematic network synthesizes the relative importance of the main eukaryotic taxonomic and functional groups³⁶ and their interactions (symbiosis sensu lato in green and predation in red). Metabarcoding (metaB) data highlighted the dominant diversity of heterotrophic and parasitic protists⁹⁵ and their central role in shaping the global plankton interactome⁹⁸. **d** | The suite of *Tara Oceans* automated imaging devices (shown here: flow-cytometry, imaging flow cytobot (IFCB), FlowCam, ZooScan and underwater vision profiler (UVP)) make possible quantification of the abundance of organisms ranging from 0.8 µm to several centimetres in size. These spectra can then be used to estimate how biomass is distributed along plankton size spectra or functional groups. eHCFM, environmental high-content fluorescence microscopy; HTI, high-throughput imaging; HTS, high-throughput sequencing; metaG, metagenomic; metaT, metatranscriptomic; SAG, single amplified genome. Part **d** adapted from REF.¹³², CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>).



► Fig. 5 | Eukaryotic shapes and symbioses explored by *Tara Oceans* plankton

imaging. All eukaryotes host a cohort of more or less specific or beneficial viral, bacterial, archaeal or eukaryotic symbionts. The staining strategy developed for automated confocal microscopy of aquatic microbial eukaryotes (environmental high-content fluorescence microscopy)⁹⁰ revealed symbiotic interactions in marine protists. **A** | Photosymbiosis occurs between the calcareous ciliate *Tiarina* sp. and *Symbiodinium* dinoflagellates. The image shows confocal laser scanning microscopy (CLSM; panels **Aa,Ab**) and scanning electron microscopy (panel **Ac**) reconstructions of the ciliate host; false colours show nuclei of the ciliate in cyan, nuclei of the symbiotic microalgae in blue and *Symbiodinium* chloroplasts in red (scale bars 20 µm). **B** | Diatoms (*Fragilariopsis* sp. cells assembled in a chain) and a heterotrophic tintinnid ciliate (*Salpingella* sp., shell with trumpet-shaped oral opening) form a symbiotic relationship (scale bars 10 µm). **C** | Intracellular cyanobacterial symbionts (*Richelia* sp.) are seen within the pennate diatom *Rhizosolenia* (panel **Ca**; scale bar 20 µm). Two cyanobacterial trichomes are visible with their heterocysts (panel **Cb**; scale bar 10 µm). **D** | Association between the heterotrophic dinoflagellate *Amphisolenia* and unidentified cyanobacteria hosted inside the cell wall (arrowhead) (scale bar 30 µm). **E** | The diatom *Corethron* sp. harbours several epiphytic nanoflagellates living in small shells and attached to the diatom cell wall (scale bar 30 µm). **F** | Dinoflagellates from the genus *Ornithocercus* host extracellular cyanobacterial symbionts in their ‘symbiotic chamber’ (OmCyn cyanobacteria)¹⁵¹ (dinoflagellate cell size between 80 and 100 µm). **G**. Giant colonial protists (*Collodaria*) harbour intracellular dinoflagellate symbionts (*Brandtodium* sp.)¹⁵². This light stereoscope image (panel **Ga**) shows an entire colony (scale bar 1 mm); the CLSM image (panel **Gb**) within a colony shows collodarian cells (blue, 200 µm), endosymbiotic dinoflagellates (red, 20–30 µm) and a reticulate cytoplasmic network (green filaments). All images (except for those in panels **Ac,Ga**) are CLSM reconstructions from single *Tara Oceans* cells. Panel **A** adapted from REF.¹¹⁶, Springer Nature Limited; panel **B** from REF.⁹⁹, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>); panels **C** and **F** from REF.¹⁵³, Springer Nature Limited; and panels **D** and **E** from REF.⁹⁰, CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>). Panel **G** provided by C.d.V.

striking regularity suggests that the assembly of protist communities is governed by large-scale ubiquitous processes, despite the highly dynamic and variable environment. The underlying drivers of this relationship are unknown, but the similarity of the power-law exponent to 3/2 resembles the temporal spectra of intermittently varying ecosystems^{113,114}, suggesting that local abundances are influenced by spatiotemporal variability. Understanding the origin and impact of this apparently universal abundance signature of non-dominant protists on plankton ecology is important for evaluating the resilience of marine biodiversity in a changing ocean.

Of note, the global eukaryotic metaB survey has revealed a realm of unknown diversity and functions among heterotrophic and symbiotic (*sensu lato*) protists (FIG. 4c). For example, planktonic diplonemids may well be the most diverse group of planktonic eukaryotes in the ocean, with the majority of their abundance and diversity in deeper waters⁹⁶. Although the underlying causes of their hyperdiversification and the roles of these different lineages in the ecosystem remain unknown, specific trophic interactions, such as bacterivory or parasitism, appear most probable^{96,115}. Sequencing of barcodes from individual protists isolated from ethanol-preserved plankton samples (Supplementary Box 3) showed widespread symbiotic associations, such as those between the coral-associated dinoflagellate *Symbiodinium* and the calcified ciliate *Tiarina*¹¹⁶ (FIG. 5A) and between the chain-forming pennate diatom *Fragilariopsis doliolus* and tintinnid ciliates⁹⁹ (FIG. 5B). Image acquisition of fragile plankton from the surface to 1,000-m depth highlighted the abundance of giant photosymbiotic rhizarian protists (order Collodaria; FIG. 5C) detected by metabarcoding in mesoplankton size

Bacterivory

Organisms that obtain carbon and energy primarily from the consumption of bacteria.

Miocene epoch

First geological epoch of the Neogene period (2.6 million to 23 million years ago) that extends from about 23 million to 5 million years ago.

fractions⁹⁵ and showed that their biomass exceeds that of all zooplankton in (sub)tropical oceans¹¹⁷.

Single-cell genomics, metaG, and metaT analyses of *Tara Oceans* eukaryote-enriched samples confirmed the hyperdiversification of heterotrophic and symbiotrophic protists and suggested potential mechanisms underlying these processes. A single-cell genomics survey revealed hidden functional complexity and niche differentiation in unculturable heterotrophic protists, partially explaining their unforeseen diversity^{93,94,118}. Although a direct comparison with bacterial and archaeal gene diversity is difficult¹⁸, global eukaryotic metaT data from 441 communities yielded an extreme richness of more than 116 million transcripts from eukaryotes (including metazoans) without apparent saturation⁴⁹. Many unknown genes were detected, and the biogeography of their specific expression revealed a potential link to niche adaptation.

On the basis of these findings, protists have arguably emerged as the group of organisms that drive today’s plankton complexity (FIG. 4a). To unify the analyses of the emerging complexity of protist data under a single ontology, an initiative for building a universal taxonomic framework for eukaryotes has been launched (UniEuk), and the *Tara Oceans* metaT data have yielded the largest available gene collection for eukaryotes⁴⁹.

Zooplankton. Zooplankton have a central role in the ocean by transferring energy, nutrients and biomass from lower to higher trophic levels^{5,119}. Biodiversity patterns in planktonic metazoans are far less understood than those in their terrestrial counterparts. In *Tara Oceans*, five different types of nets (Supplementary Table 3) were used to collect nearly 1,500 standardized zooplankton samples at depths from the surface to a few hundred meters. Imaging and HTS were then used (FIG. 4b) to assess the morphogenetic complexity of zooplankton communities in well-defined oceanographic provinces.

In terms of imaging, all zooplankton samples have been processed, and most have been validated by experts using EcoTaxa (Supplementary Table 2). Several *Tara Oceans* imaging data sets were used to assess the mechanisms that contribute to the limited dispersal of Indian Ocean plankton populations into the Atlantic⁵². Imaging data have also demonstrated substantially reduced abundances of metazoan plankton in the Indian Ocean oxygen minimum zone and its positive effect on carbon flux¹²⁰.

Analyses were also performed on targeted zooplankton taxa. For example, combined morphological and phylogenetic data revealed that sea snails (clade Thecosomata) diversified through four major morphogenetic events that coincide with climate events from the Eocene epoch to the Miocene epoch; this evolutionary scenario is potentially driven by skeleton selection to avoid predation or to increase buoyancy¹²¹. Additionally, a comprehensive phylogenetic study of chaetognaths based on complete ribosomal DNA genes amplified from preserved specimens showed that their evolution corresponds to simplification of a pre-existing body plan rather than to an increase in morphological complexity¹²².

Finally, metaG data have been used to study the population structure of the abundant copepod *Oithona* in the Mediterranean Sea¹²³, providing evidence for genes

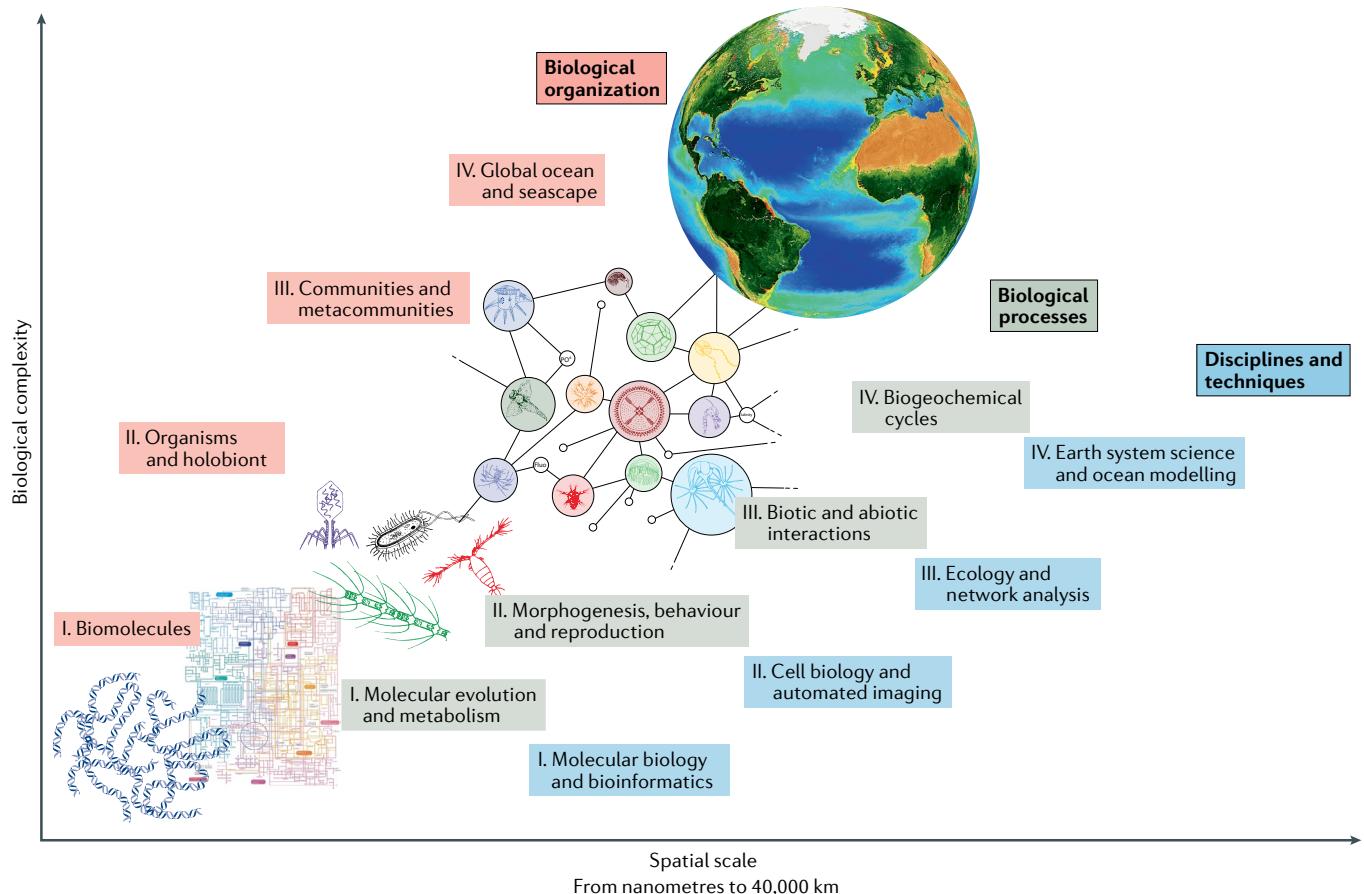


Fig. 6 | Ecosystems biology and integrative analyses of the global oceans. The planetary-scale and large volumes of the *Tara Oceans* data sets for the epipelagic and mesopelagic ocean allow the extraction of emerging properties that one can successively integrate into higher levels of biological organization. This step-by-step simplification and integration of complexity can be compared with Russian dolls in providing an eco-evolutionary, data-driven framework for modelling of the Earth system. The analysis of each layer requires the use of different techniques, which are often discipline specific, to target different biological processes. Besides the interest within each layer of spatial scale and biological complexity (described from I to IV), *Tara Oceans* has integrated information across the various layers of biological organization. For example, one study⁹⁸ built a co-occurrence network (III) using organismal abundance data (II) inferred from metabarcoding data (I) and validated examples using imaging of organisms (II). Another study⁵² analysed plankton communities (III) in the context of oceanographic models (IV) using DNA sequence data (I). Other studies have used the whole ecosystems biology framework (I to IV) to scale up from genes and organisms to emergent ocean ecosystem processes such as the biological carbon pump^{45,125}.

under selection in specific contexts and allowing the creation of a collection of single-nucleotide polymorphisms in a reference-independent manner¹²⁴. The current data pave the way for studies of diversity and expression at the gene level for the main groups of zooplankton; however, these studies merely scratch the surface of the morphogenetic information buried in the *Tara Oceans* collection of zooplankton samples and data. Increased efforts to sequence metaB data and genomes of the major organisms, to use metaG and metaT information for detecting genes under active selection and to correlate genetic data to imaging information in the future will undoubtedly advance our knowledge of these important planktonic organisms and their role in the ocean ecosystem.

Integrative ocean ecosystems biology

A unique feature of the growing *Tara Oceans* data set is its relatively uniform and deep coverage over spatial and taxonomic scales, encompassing the variability of

the global plankton ecosystem from the surface to mesopelagic depths (FIGS 1,2). This scope and the large data sets derived from it facilitate data-driven analyses to extract information in a comprehensive eco-evolutionary framework. *Tara Oceans* thus attempted to integrate the different layers of ecosystem organization, from genes to organismal populations, across environmental and spatial variations and beyond analyses of plankton within specific size fractions (FIG. 6).

To decipher the plankton metacommunity structure, a global plankton co-occurrence network was drafted to include both biotic and abiotic information⁹⁸. The results showed that biotic and positive co-occurrences predominate over environmental influences on community structure. Furthermore, this network revealed the prevalence of parasitic and photosymbiotic protists^{95,117} as keystone taxa that increase the connectivity of plankton food webs (FIG. 4C). Together, plankton metabarcoding⁹⁵, 3D fluorescence microscopy^{90,99,116} and underwater

Agulhas choke point
Oceanic system south of Africa where warm and salty Indian Ocean waters leak into the South Atlantic Ocean impacting the global oceanic circulation.

imaging¹¹⁷ confirmed that putative symbioses, including parasitism and mutualism, derived from global co-occurrences are ubiquitous in plankton ecology (FIG. 5) and may underlie the hyperdiversification of protists⁹⁵ through evolutionary mechanisms that remain elusive. Overall, this notion challenges traditional ocean ecology that focuses on the negative relationships between producers and consumers and most of the models that typically ignore symbioses to predict nutrient and energy flow in the ocean.

Tara Oceans data were used to address basin-scale oceanographic questions in a study on the impact of the Agulhas choke point on plankton communities⁵². Specifically, connectivity between the Indian Ocean and the Atlantic Ocean influenced different planktonic groups in different ways, largely as a function of their size. The Agulhas rings were important conduits for transporting plankton between the two oceans. These findings highlight the need to investigate the relationship between plankton diversity and global ocean circulation. Specifically, the Agulhas current retroflection was a key factor constraining diatom diversity, in line with previous palaeo-oceanographic studies based on diatom microfossils^{100,112}.

Additionally, *Tara* Oceans data have been used in graph-based methods from systems biology to integrate the full suite of ecological, morphological and genetic information for inclusion of biological complexity in ocean modelling (FIG. 6). A study using network-partitioning⁴⁵ identified plankton subcommunities and gene modules associated with carbon export from the upper epipelagic zone to the ocean interior, demonstrating the possibility to scale up from genes to ecosystems and to derive insightful models of key ocean biogeochemical processes. In addition to viruses emerging as the best predictors for the variability of carbon export in the oligotrophic ocean⁴⁵, the same graph-based methods showed that plankton subcommunities varied with the iron products from two global-scale biogeochemical models¹²⁵. Within these subcommunities, genomic adaptation based on gene-copy numbers was disentangled from transcriptomic adaptation based on gene expression for specific groups and functions. For example, many photosynthetic protists respond to iron limitation by shifting the use of a key gene coding for ferredoxin to an iron-independent analogue, flavodoxin. The rapidly responding groups, such as diatoms, are frequently adapted at the genomic level by harbouring variable numbers of each gene depending on optimal growth conditions. In contrast, other organisms, such as haptophytes or pelagophytes, rely on differential transcription to shift to the best analogue^{49,125}. Such meta-omics analyses were used to explore the underpinnings of recurrent phytoplankton blooms in the Marquesas archipelago in the central equatorial Pacific Ocean, and revealed that an increase in iron bioavailability is likely to be the underlying cause of the blooms¹²⁵. This example demonstrates that the field of ocean meta-omics is now sufficiently mature to provide an independent, biologically based validation of ecosystem models. In another case, the abundance and expression of transporter genes in diatoms was determined as a function of environmental variation, and the observed variation was then used to train an algorithm to predict the functional response of

diatoms to future seawater temperatures¹²⁶. The combination of global biogeochemical models with genomics and community composition analysis highlights the transformative nature of integrating quantitative omics data and oceanography to better understand the functions of marine ecosystems¹²⁷.

More recently, latitudinal gradients and global predictors of plankton diversity across archaea, bacteria, eukaryotes and major virus clades have been explored with use of molecular and imaging data from *Tara* Oceans³⁶. Latitudinal diversity gradients were previously studied primarily in terrestrial macroorganisms and typically consist of a monotonic poleward decline of local diversity¹²⁸. Studies in ocean ecosystems have been fragmentary and have often led to different results^{129,130}; thus, the availability of a single comprehensive data set that represents all planktonic organisms collected on a global scale made possible investigations of such macroecological patterns. There was a decline of diversity for most planktonic groups towards the poles, and this decline was mainly driven by temperature with input from productivity and seasonal variability³⁶. Projections into the future using climate models of the Intergovernmental Panel on Climate Change further suggested that severe warming of the ocean in the future may lead to tropicalization of the diversity of most planktonic groups at higher latitudes. These changes may have ripple effects on marine ecosystem functioning, affecting both biogeochemical cycles and trophic interactions globally.

Overall, these ecosystems biology approaches highlighted previously underrated organisms and genes that should be assessed as genomic proxies for the prediction of key emergent ecosystem functions. These analyses were unprecedented with respect to the global ecosystem scale and further laid the foundation for future robust ecosystem modelling to bridge information about genes, organisms, consortia and biomes (FIG. 6).

Conclusions and perspectives

Life has evolved over billions of years, starting in the oceans; however, it is only recently that technologies have enabled us to capture the taxonomic, genetic and morphological biodiversity of extant ocean life as a whole, from microorganisms to animals. *Tara* Oceans exemplifies how such a holistic approach has been used to study ocean plankton at a planetary scale. Starting from an adventurous initiative of blue-sky research with no substantial core funding, the project has developed into a multinational, multidisciplinary, collaborative programme (BOX 1). The comprehensive end-to-end sampling protocols developed to capture plankton from viruses to metazoans, organisms rarely studied together, have greatly expanded our knowledge of biodiversity, organismal interactions, ecological drivers of community structure and genomic proxies for key ecosystem processes, such as carbon export, in the ocean. This approach has already prompted similar implementations on other oceanographic cruises and time-series studies, and may, ultimately, help to establish much-needed standards for biological sampling in oceanography^{131–133}. Furthermore, the commitment to create a consistent knowledge

base has resulted in open-access resources of in situ multi-omics sequencing, imaging and environmental data, which a diverse community of researchers has been mining ever since for new insights and discoveries.

Although *Tara* Oceans maximized its global reach in its sampling design, there remains a need to increase the geographic coverage and granularity of ocean sampling, also across depth. In the meantime, additional *Tara* expeditions have sampled transects across parts of the North Atlantic Ocean and Pacific Ocean¹³⁴ including coastal reef waters¹³⁵. However, the subarctic North Pacific, equatorial sections and the Southern Ocean remain priority areas with insufficient coverage. In addition to complementary ocean sampling campaigns (for example, Ocean Sampling Day¹⁵ and the International Census of Marine Microbes¹⁶), repeated cruises and expeditions applying similar approaches have been completed, are under way or are planned¹³⁶, and will help to close these gaps. One limitation inherent to the spatially distributed nature of ocean sampling expeditions is the lack of temporal resolution. To complement existing snapshots of planktonic states, it will thus be important to incorporate trajectories of community variability over time as they have been recorded, for example, at long-term ocean time-series stations^{12,14,137–141}, in shorter-term studies at day-to-day resolution¹⁴² and during mesoscale process studies¹⁴³. To further increase spatiotemporal information of plankton dynamics, these local measurements will ideally be complemented by future technological advances¹⁴⁴ to provide global and multiyear coverage, for example using in situ remote observatories for automated genomic, imaging and environmental data collection¹⁴⁵.

In conclusion, considering that ocean ecosystems biology aims to gain a holistic understanding of the biodiversity and processes that govern the ocean, the field is still very much in a data-driven, phenomenological

discovery phase¹⁴⁶. And yet it must rapidly get up to speed as anthropogenic climate change is already altering the global ocean¹⁴⁷. Moreover, ocean plankton will not only be affected by the outcomes of these changes but will also, to some degree, control them¹⁴⁸. To effectively evaluate how open-ocean life will respond to environmental change, large-scale, diverse, interdisciplinary efforts — including empirical, theoretical and modelling approaches — are needed to advance our understanding of organismal abundances and biomass, physiology and interactions across space and time. To address this need, members of research teams need to step far outside their comfort zone, and non-traditional funding schemes will be required to support the syntheses needed to make transformative advances in such a complex space. Such scientific endeavours and the resulting information must be coupled with concerted efforts to inform policy and management decisions, and to provide diverse outreach programmes, including technology transfer and capacity building, to truly foster a societal impact on Earth (BOX 2).

Furthermore, the ocean biome is intricately linked to other biomes on Earth, including host-associated systems. The crucial need for an integrated understanding of ecosystem processes across the ocean, atmosphere and terrestrial systems could be fostered by the availability of discoverable, accessible, interoperable and reusable data, and toolkits that would facilitate global-scale synthetic analyses, similarly to what is already in practice in ongoing international efforts in the areas of physical oceanography (for example, Argo programme, NASA) and the health sector (for example, Global Alliance for Genomics and Health, International Cancer Genome Consortium). Together, such efforts should ultimately help us to better understand and predict the effect of climate change on extant life and the future habitability of our planet in the Anthropocene epoch.

Published online 12 May 2020

- Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
- Guidi, L. et al. A new look at ocean carbon remineralization for estimating deepwater sequestration. *Global Biogeochem. Cycles* **29**, 1044–1059 (2015).
- Henson, S. A., Sanders, R. & Madsen, E. Global patterns in efficiency of particulate organic carbon export and transfer to the deep ocean. *Global Biogeochem. Cycles* **26**, GB1028 (2012).
- Kwon, E. Y., Primeau, F. & Sarmiento, J. L. The impact of remineralization depth on the air-sea carbon balance. *Nat. Geosci.* **2**, 630–635 (2009).
- Azam, F. et al. The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* **10**, 257–263 (1983).
- Raes, J. & Bork, P. Molecular eco-systems biology: towards an understanding of community function. *Nat. Rev. Microbiol.* **6**, 693–699 (2008).
- Karsenti, E. et al. A holistic approach to marine eco-systems biology. *PLoS Biol.* **9**, e1001177 (2011).
- Rusch, D. B. et al. The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
- Venter, J. C. et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
- This study applies high-throughput DNA sequencing to produce a large data set of microbial community genome fragments from surface seawaters of the Sargasso Sea and identifies more than 1.2 million previously unknown genes, illustrating the diversity of ocean microbial life.**
- Yooseph, S. et al. The Sorcerer II global ocean sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5**, e16 (2007).
- Pesant, S. et al. Open science resources for the discovery and analysis of *Tara* Oceans data. *Sci. Data* **2**, 150023 (2015).
- Biller, S. J. et al. Marine microbial metagenomes sampled across space and time. *Sci. Data* **5**, 180176 (2018).
- Duarte, C. M. Seafaring in the 21st century: the Malaspina 2010 Circumnavigation Expedition. *Limnol. Oceanogr. Bull.* **24**, 11–14 (2015).
- Karl, D. M. & Church, M. J. Microbial oceanography and the Hawaii ocean time-series programme. *Nat. Rev. Microbiol.* **12**, 699–713 (2014).
- Kopf, A. et al. The ocean sampling day consortium. *Gigascience* **4**, 27 (2015).
- Amaral-Zettler, L. et al. in *Life in the World's Oceans* (ed. McIntyre, A. D.) 221–245 (Wiley, 2010).
- Longhurst, A. Seasonal cycles of pelagic production and consumption. *Prog. Oceanogr.* **36**, 77–167 (1995).
- Sunagawa, S., Karsenti, E., Bowler, C. & Bork, P. Computational eco-systems biology in *Tara* Oceans: translating data into knowledge. *Mol. Syst. Biol.* **11**, 809 (2015).
- Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541–548 (1999).
- Suttle, C. A. Marine viruses-major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
- Wommack, K. E. & Colwell, R. R. Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000).
- Brum, J. R. & Sullivan, M. B. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* **13**, 147–159 (2015).
- Brum, J. R. et al. Patterns and ecological drivers of ocean viral communities. *Science* **348**, 1261498 (2015).
- This article describes the first of the *Tara* Oceans efforts to investigate the diversity and structure of double-stranded DNA viral communities in the oceans, supporting a model of passive global transport by ocean currents and selection by local environmental conditions.**
- Gregory, A. C. et al. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **177**, 1109–1123.e14 (2019).
- Roux, S. et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**, 689–693 (2016).
- Duhaime, M. B. et al. Comparative omics and trait analyses of marine *pseudoalteromonas* phages advance the phage OTU concept. *Front. Microbiol.* **8**, 1241 (2017).
- Martinez-Hernandez, F. et al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat. Commun.* **8**, 15892 (2017).
- Warwick-Dugdale, J. et al. Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* **7**, e6800 (2019).
- Nishimura, Y. et al. Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere* **2**, e00359-16 (2017).
- Nishimura, Y. et al. ViPTree: the viral proteomic tree server. *Bioinformatics* **33**, 2379–2380 (2017).

31. Bin Jang, H. et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).
32. Bolduc, B. et al. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and bacteria. *PeerJ* **5**, e3243 (2017).
33. Roux, S. et al. Minimum information about an uncultivated virus genome (MIUViG). *Nat. Biotechnol.* **37**, 29–37 (2019).
34. Simmonds, P. et al. Consensus statement: virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* **15**, 161–168 (2017).
35. Baas-Becking, L. G. M. *Geobiologie of Inleiding tot de Milieukunde* (Van Stockum & Zoon, 1934).
36. Ibarbalz, F. M. et al. Global trends in marine plankton diversity across kingdoms of life. *Cell* **179**, 1084–1097.e21 (2019).
37. Jia, Y., Shan, J., Millard, A., Clokie, M. R. & Mann, N. H. Light-dependent adsorption of photosynthetic cyanophages to *Synechococcus* sp. WH7803. *FEMS Microbiol. Lett.* **310**, 120–126 (2010).
38. Ribalet, F. et al. Light-driven synchrony of Prochlorococcus growth and mortality in the subtropical Pacific gyre. *Proc. Natl Acad. Sci. USA* **112**, 8008–8012 (2015).
39. Yoshida, T. et al. Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan. *ISME J.* **12**, 1287–1295 (2018).
40. Fridman, S. et al. A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected Prochlorococcus cells. *Nat. Microbiol.* **2**, 1350–1357 (2017).
41. Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* **424**, 741 (2003).
42. Sullivan, M. B. et al. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* **4**, e234 (2006).
43. Hurwitz, B. L., Hallam, S. J. & Sullivan, M. B. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol.* **14**, R123 (2013).
44. Howard-Varona, C. et al. Regulation of infection efficiency in a globally abundant marine Bacterioidetes virus. *ISME J.* **11**, 284–295 (2017).
45. Guidi, L. et al. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**, 465–470 (2016). **This study integrates Tara Oceans data across organismal size classes from epipelagic depths, revealing that unexpected taxa can predict the downward export of carbon by biological processes in subtropical, nutrient-depleted oceans.**
46. Howard-Varona, C. et al. Phage-specific metabolic reprogramming of virocells. *ISME J.* **14**, 881–895 (2020).
47. Wilhelm, S. W. & Suttle, C. A. Viruses and nutrient cycles in the sea – viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* **49**, 781–788 (1999).
48. Hurwitz, B. L., Brum, J. R. & Sullivan, M. B. Depth-stratified functional and taxonomic niche specialization in the ‘core’ and ‘flexible’ Pacific Ocean virome. *ISME J.* **9**, 472–484 (2015).
49. Carradec, Q. et al. A global ocean atlas of eukaryotic genes. *Nat. Commun.* **9**, 373 (2018).
50. Hingamp, P. et al. Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME J.* **7**, 1678–1695 (2013).
51. Lescot, M. et al. Reverse transcriptase genes are highly abundant and transcriptionally active in marine plankton assemblages. *ISME J.* **10**, 1134–1146 (2016).
52. Villar, E. et al. Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science* **348**, 1261447 (2015).
53. Li, Y. et al. The earth is small for “Leviathans”: long distance dispersal of giant viruses across aquatic environments. *Microbes Environ.* **34**, 334–339 (2019).
54. Mihara, T. et al. Taxon richness of “Megaviridae” exceeds those of bacteria and archaea in the ocean. *Microbes Environ.* **33**, 162–171 (2018).
55. 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).
56. Clerissi, C. et al. Deep sequencing of amplified Prasinovirus and host green algal genes from an Indian Ocean transect reveals interacting trophic dependencies and new genotypes. *Environ. Microbiol. Rep.* **7**, 979–989 (2015).
57. Clerissi, C. et al. Unveiling of the diversity of prasinoviruses (Phycodnaviridae) in marine samples by using high-throughput sequencing analyses of PCR-amplified DNA polymerase and major capsid protein genes. *Appl. Environ. Microbiol.* **80**, 3150–3160 (2014).
58. Clerissi, C. et al. Prasinovirus distribution in the northwest Mediterranean Sea is affected by the environment and particularly by phosphate availability. *Virology* **466–467**, 146–157 (2014).
59. Li, Y. et al. Degenerate PCR primers to reveal the diversity of giant viruses in coastal waters. *Viruses* **10** (2018).
60. Blanc-Mathieu, R. et al. Viruses of the eukaryotic plankton are predicted to increase carbon export efficiency in the global sunlit ocean. Preprint at *bioRxiv* <https://doi.org/10.1101/710228> (2019).
61. Bolduc, B., Youens-Clark, K., Roux, S., Hurwitz, B. L. & Sullivan, M. B. iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. *ISME J.* **11**, 7–14 (2017).
62. Miura, T. et al. Linking virus genomes with host taxonomy. *Viruses* **8**, 66 (2016).
63. Steward, G. F. et al. Are we missing half of the viruses in the ocean? *ISME J.* **7**, 672–679 (2013).
64. Salazar, G. et al. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* **179**, 1068–1083.e21 (2019).
65. Sunagawa, S. et al. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015). **This study catalogues 40 million ocean microbial genes and shows temperature to be a main driver of open-ocean microbial community composition in the epipelagic zone at a global scale.**
66. Kultima, J. R. et al. MOCAT: a metagenomics assembly and gene prediction toolkit. *PLoS One* **7**, e47656 (2012).
67. Li, J. et al. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
68. Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
69. DeLong, E. F. et al. Community genomics among stratified microbial assemblages in the ocean’s interior. *Science* **311**, 496–503 (2006).
70. Giovannoni, S. J. & Stingl, U. Molecular diversity and ecology of microbial plankton. *Nature* **437**, 343–348 (2005).
71. Fuhrman, J. A. et al. Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc. Natl Acad. Sci. USA* **103**, 13104–13109 (2006).
72. Milanese, A. et al. Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat. Commun.* **10**, 1014 (2019).
73. Farrant, G. K. et al. Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc. Natl Acad. Sci. USA* **113**, E3365–E3374 (2016).
74. Greber, T. et al. Light color acclimation is a key process in the global ocean distribution of *Synechococcus* cyanobacteria. *Proc. Natl Acad. Sci. USA* **115**, E2010–E2019 (2018).
75. Yelton, A. P. et al. Global genetic capacity for mixotrophy in marine picocyanobacteria. *ISME J.* **10**, 2946–2957 (2016).
76. Royo-Llonch, M. et al. Exploring microdiversity in novel *Kordia* sp. (Bacteroidetes) with proteorhodopsin from the tropical Indian Ocean via single amplified genomes. *Front. Microbiol.* **8**, 1317 (2017).
77. Royo-Llonch, M., Sánchez, P., González, J. M., Pedrós-Alio, C. & Acinas, S. G. Ecological and functional capabilities of an uncultured *Kordia* sp. *Syst. Appl. Microbiol.* **43**, 126045 (2020).
78. Cabello, A. M. et al. Global distribution and vertical patterns of a prymnesiophyte-cyanobacteria obligate symbiosis. *ISME J.* **10**, 693–706 (2016).
79. Cornejo-Castillo, F. M. et al. Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nat. Commun.* **7**, 11071 (2016).
80. Cornejo-Castillo, F. M. et al. UCYN-A3, a newly characterized open ocean sublineage of the symbiotic N2-fixing cyanobacterium *Candidatus Atelocyanobacterium thalassa*. *Environ. Microbiol.* **21**, 111–124 (2019).
81. Delmont, T. O. et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* **3**, 804–813 (2018).
82. Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T. J. G. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557**, 101–105 (2018). **This study exemplifies the use of Tara Oceans data to formulate new hypotheses by reconstructing genomes that support a mitochondrial origin before the divergence of all Alphaproteobacteria sampled to date.**
83. Parks, D. H. et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542 (2017).
84. Tully, B. J., Graham, E. D. & Heidelberg, J. F. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci. Data* **5**, 170203 (2018).
85. Pushkarev, A. et al. A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature* **558**, 595–599 (2018).
86. Oppermann, J. et al. MerMAIDs: a family of metagenomically discovered marine anion-conducting and intensely desensitizing channelrhodopsins. *Nat. Commun.* **10**, 3315 (2019).
87. Louca, S., Parfrey, L. W. & Doebeli, M. Decoupling function and taxonomy in the global ocean microbiome. *Science* **353**, 1272–1277 (2016).
88. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc. Natl Acad. Sci. USA* **115**, 6506–6511 (2018).
89. Caron, D. A., Countway, P. D., Jones, A. C., Kim, D. Y. & Schnetzer, A. Marine protistan diversity. *Ann. Rev. Mar. Sci.* **4**, 467–493 (2012).
90. Colin, S. et al. Quantitative 3D-imaging for cell biology and ecology of environmental microbial eukaryotes. *eLife* **6**, e26066 (2017).
91. Decelle, J. et al. PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. *Mol. Ecol. Resour.* **15**, 1435–1445 (2015).
92. Guillou, L. 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).
93. Seeleuthner, Y. et al. Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. *Nat. Commun.* **9**, 310 (2018).
94. Sieracki, M. E. et al. Single cell genomics yields a wide diversity of small planktonic protists across major ocean ecosystems. *Sci. Rep.* **9**, 6025 (2019).
95. de Vargas, C. et al. Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605 (2015). **This study surveys the eukaryotic diversity of ocean plankton from the smallest protists to millimetre-sized animals by 18S ribosomal RNA gene amplicon sequencing, revealing 150,000 taxonomic groups dominated by protistan parasites and symbiotic hosts.**
96. Flegontova, O. et al. Extreme diversity of diplomonad eukaryotes in the ocean. *Curr. Biol.* **26**, 3060–3065 (2016).
97. Decelle, J. et al. Worldwide occurrence and activity of the reef-building coral symbiont *Symbiodinium* in the open ocean. *Curr. Biol.* **28**, 3625–3633.e3623 (2018).
98. Lima-Mendez, G. et al. Determinants of community structure in the global plankton interactome. *Science* **348**, 1262073 (2015). **This study evaluates the effect of abiotic and biotic factors on organismal interactions among bacteria, archaea, eukaryotes and viruses, emphasizing the role of grazing, pathogenicity and parasitism as predictors of plankton community structure.**
99. Vincent, F. J. et al. The epibiotic life of the cosmopolitan diatom *Fragilariaopsis doliolus* on heterotrophic ciliates in the open ocean. *ISME J.* **12**, 1094–1108 (2018).
100. Malviya, S. et al. Insights into global diatom distribution and diversity in the world’s ocean. *Proc. Natl Acad. Sci. USA* **113**, E1516–E1525 (2016).
101. Le Bescot, N. et al. Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding. *Environ. Microbiol.* **18**, 609–626 (2016).
102. Lopes Dos Santos, A. et al. Diversity and oceanic distribution of prasinophytes clade VII, the dominant group of green algae in oceanic waters. *ISME J.* **11**, 512–528 (2017).

103. Gimmerer, A., Korn, R., de Vargas, C., Audic, S. & Stoeck, T. The *Tara Oceans* voyage reveals global diversity and distribution patterns of marine planktonic ciliates. *Sci. Rep.* **6**, 33555 (2016).
104. Beaugrand, G., Luczak, C., Goerville, E. & Kirby, R. R. Marine biodiversity and the chessboard of life. *PLoS One* **13**, e0194006 (2018).
105. Biard, T. et al. Biogeography and diversity of Collodaria (Radiolaria) in the global ocean. *ISME J.* **11**, 1331–1344 (2017).
106. Del Campo, J. et al. Assessing the diversity and distribution of apicomplexans in host and free-living environments using high-throughput amplicon data and a phylogenetically informed reference framework. *Front. Microbiol.* **10**, 2373 (2019).
107. Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **11**, 2639–2643 (2017).
108. Foster, Z. S., Sharpton, T. J. & Grunwald, N. J. Metacoder: an R package for visualization and manipulation of community taxonomic diversity data. *PLoS Comput. Biol.* **13**, e1005404 (2017).
109. Pierella Karlusich, J. J., Ibarbalz, F. M. & Bowler, C. Phytoplankton in the *Tara Ocean*. *Annu. Rev. Mar. Sci.* **12**, 233–265 (2020).
110. Leblanc, K. et al. Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export. *Nat. Commun.* **9**, 953 (2018).
111. Treguer, P. et al. Influence of diatom diversity on the ocean biological carbon pump. *Nat. Geosci.* **11**, 27–37 (2018).
112. Rabosky, D. L. & Sorhannus, U. Diversity dynamics of marine planktonic diatoms across the Cenozoic. *Nature* **457**, 183–186 (2009).
113. Azalea, S., Pigolotti, S., Banavar, J. R. & Maritan, A. Dynamical evolution of ecosystems. *Nature* **444**, 926–928 (2006).
114. Ferriere, R. & Cazelles, B. Universal power laws govern intermittent rarity in communities of interacting species. *Ecology* **80**, 1505–1521 (1999).
115. Gawryluk, R. M. R. et al. Morphological identification and single-cell genomics of marine diplomonads. *Curr. Biol.* **26**, 3053–3059 (2016).
116. Mordret, S. et al. The symbiotic life of Symbiodinium in the open ocean within a new species of calcifying ciliate (*Tiarina* sp.). *ISME J.* **10**, 1424–1436 (2016).
117. Biard, T. et al. In situ imaging reveals the biomass of giant protists in the global ocean. *Nature* **532**, 504–507 (2016).
118. Vannier, T. et al. Survey of the green picoalga *Bathycoccus* genomes in the global ocean. *Sci. Rep.* **6**, 37900 (2016).
119. Steinberg, D. K. & Landry, M. R. Zooplankton and the ocean carbon cycle. *Ann. Rev. Mar. Sci.* **9**, 413–444 (2017).
120. Roullier, F. et al. Particle size distribution and estimated carbon flux across the Arabian Sea oxygen minimum zone. *Biogeosciences* **11**, 4541–4557 (2014).
121. Corse, E. et al. Phylogenetic analysis of Thecosomatidae (holoplanktonic opisthobranchia) using morphological and molecular data. *PLoS One* **8**, e59439 (2013).
122. Gasmi, S. et al. Evolutionary history of Chaetognatha inferred from molecular and morphological data: a case study for body plan simplification. *Front. Zool.* **11**, 84 (2014).
123. Madou, M. A. et al. New insights into global biogeography, population structure and natural selection from the genome of the epipelagic copepod *Oithona*. *Mol. Ecol.* **26**, 4467–4482 (2017).
124. Arif, M. et al. Discovering millions of plankton genomic markers from the Atlantic Ocean and the Mediterranean Sea. *Mol. Ecol. Resour.* **19**, 526–535 (2019).
125. Caputi, L. et al. Community-level responses to iron availability in open ocean plankton ecosystems. *Global Biogeochem. Cycles* **33**, 391–419 (2019).
126. Busseni, G. et al. Meta-omics reveals genetic flexibility of diatom nitrogen transporters in response to environmental changes. *Mol. Biol. Evol.* **36**, 2522–2535 (2019).
127. D'Alelio, D. et al. Modelling the complexity of plankton communities exploiting omics potential: From present challenges to an integrative pipeline. *Curr. Opin. Syst. Biol.* **13**, 68–74 (2019).
128. Whittaker, R. H. Evolution and measurement of species diversity. *Taxon* **21**, 213–251 (1972).
129. Fuhrman, J. A. et al. A latitudinal diversity gradient in planktonic marine bacteria. *Proc. Natl Acad. Sci. USA* **105**, 7774–7778 (2008).
130. Raes, E. J. et al. Oceanographic boundaries constrain microbial diversity gradients in the South Pacific Ocean. *Proc. Natl Acad. Sci. USA* **115**, E8266–E8275 (2018).
131. Capotondi, A. et al. Observational needs supporting marine ecosystems modeling and forecasting: from the global ocean to regional and coastal systems. *Front. Mar. Sci.* **6**, 623 (2019).
132. Lombard, F. et al. Globally consistent quantitative observations of planktonic ecosystems. *Front. Mar. Sci.* **6**, 196 (2019).
133. Ten Hoopen, P. et al. Marine microbial biodiversity, bioinformatics and biotechnology (M2B3) data reporting and service standards. *Stand. Genomic Sci.* **10**, 20 (2015).
134. Gorsky, G. et al. Expanding *Tara Oceans* protocols for underway, ecosystemic sampling of the ocean-atmosphere interface during *Tara Pacific* expedition (2016–2018). *Front. Mar. Sci.* **6**, 750 (2019).
135. Planes, S. et al. The *Tara Pacific* expedition — a pan-ecosystemic approach of the “omics” complexity of coral reef holobionts across the Pacific Ocean. *PLoS Biol.* **17**, e3000483 (2019).
136. Bolhuis, H. et al. Atlantic Ocean Research Alliance — marine microbiome roadmap (AORA, 2020).
137. Cram, J. A. et al. Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *ISME J.* **9**, 563–580 (2015).
138. D'Alcalà, M. R. et al. Seasonal patterns in plankton communities in a pluriannual time series at a coastal Mediterranean site (Gulf of Naples): an attempt to discern recurrences and trends. *Sci. Mar.* **68**, 65–83 (2004).
139. Gilbert, J. A. et al. The taxonomic and functional diversity of microbes at a temperate coastal site: a ‘multi-omic’ study of seasonal and diel temporal variation. *PLoS One* **5**, e15545 (2010).
140. Romagnan, J. B. et al. Comprehensive model of annual plankton succession based on the whole-plankton time series approach. *PLoS One* **10**, e0119219 (2015).
141. Gasol, J. M. et al. ICES phytoplankton and microbial plankton status report 2009/2010 (eds O'Brien, T. D., Li, W. K. W. & Morán, X. A. G.) 138–141 (ICES, 2012).
142. Martin-Platero, A. M. et al. High resolution time series reveals cohesive but short-lived communities in coastal plankton. *Nat. Commun.* **9**, 266 (2018).
143. Laber, C. P. et al. Coccilithovirus facilitation of carbon export in the North Atlantic. *Nat. Microbiol.* **3**, 537–547 (2018).
144. Marx, V. When microbiologists plunge into the ocean. *Nat. Methods* **17**, 133–136 (2020).
145. Buttigieg, P. L. et al. Marine microbes in 4D-using time series observation to assess the dynamics of the ocean microbiome and its links to ocean health. *Curr. Opin. Microbiol.* **43**, 169–185 (2018).
146. Shneider, A. M. Four stages of a scientific discipline: four types of scientist. *Trends Biochem. Sci.* **34**, 217–223 (2009).
147. Karl, D. M. A sea of change: biogeochemical variability in the North Pacific Subtropical Gyre. *Ecosystems* **2**, 181–214 (1999).
148. Cavichioli, R. et al. Scientists’ warning to humanity: microorganisms and climate change. *Nat. Rev. Microbiol.* **17**, 569–586 (2019).
- This review article provides a consensus statement, the ‘microbiologists’ warning to humanity’, documenting how microorganisms will affect and will be affected by climate change.**
149. Bork, P. et al. *Tara Oceans* studies plankton at planetary scale. *Introduction. Science* **348**, 873 (2015).
150. Logares, R. et al. Metagenomic 16S rRNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ. Microbiol.* **16**, 2659–2671 (2014).
151. Nakayama, T. et al. Single-cell genomics unveiled a cryptic cyanobacterial lineage with a worldwide distribution hidden by a dinoflagellate host. *Proc. Natl Acad. Sci. USA* **116**, 15973–15978 (2019).
152. Probert, I. et al. Brandtiodinium gen. nov. and B. nutricula comb. Nov. (Dinophyceae), a dinoflagellate commonly found in symbiosis with polycystine radiolarians. *J. Phycol.* **50**, 388–399 (2014).
153. Decelle, J., Colin, S. & Foster, R. A. in *Marine Protists: Diversity and Dynamics* (eds Ohtsuka, S. et al.) 465–500 (Springer, 2015).
- Acknowledgements**
Tara Oceans (which includes the *Tara Oceans* and *Tara Oceans* Polar Circle expeditions) would not exist without the leadership of the Tara Ocean Foundation and the continuous support of 23 institutes (<https://oceans.taraexpeditions.org/>). The authors further thank the commitment of the following sponsors: the French CNRS (in particular Groupement de Recherche GDR3280 and the Research Federation for the Study of Global Ocean Systems Ecology and Evolution FR2022/*Tara GOSEE*), the French Facility for Global Environment (FFEM), the European Molecular Biology Laboratory, Genoscope/CEA, the French Ministry of Research and the French Government Investissements d'Avenir programmes OCEANOMICS (ANR-11-BTBR-0008), FRANCE GENOMIQUE (ANR-10-INBS-09-08) and MEMO LIFE (ANR-10-LABX-54), the PSL research university (ANR-11-IDEX-0001-02) and EMBRC-France (ANR-10-INBS-02). Funding for the collection and processing of the *Tara Oceans* data set was provided by the NASA Ocean Biology and Biogeochemistry Program under grants NNX11AQ14C, NNX09AU43G, NNX13AE58G and NNX15AC08G (to the University of Maine), the Canada Excellence Research Chair in Remote Sensing of Canada's New Arctic Frontier and the Canada Foundation for Innovation. The authors also thank agnès b. and E. Bourgois, the Prince Albert II de Monaco Foundation, the Veolia Foundation, Region Bretagne, Lorient Agglomeration, Serge Ferrari, Worldcourier and KAUST for support and commitment. The global sampling effort was made possible by countless scientists and crew who performed sampling aboard the *Tara* from 2009 to 2013, and the authors thank MERCATOR-CORIOLIS and ACRI-ST for providing daily satellite data during the expeditions. The authors are also grateful to the countries that graciously granted sampling permission. The authors thank N. Le Bescot and N. Henry for their help in designing the figures in this article. C.d.V. thanks the Roscoff Bioinformatics platform ABiMS (<http://abims.sbg-roscos.fr/>). S. Sunagawa thanks the European Molecular Biology Laboratory and ETH Zürich's high-performance computing facilities for computational support. C.B. acknowledges funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement 835067) as well as the Radcliffe Institute of Advanced Study at Harvard University for a scholar's fellowship during the 2016–2017 academic year. M.B.S. thanks the Gordon and Betty Moore Foundation (award 3790) and the US National Science Foundation (awards OCE#1536989 and OCE#1829831) as well as the Ohio Supercomputer for computational support. S.G.A. thanks the Spanish Ministry of Economy and Competitiveness (CTM2017-87736-R). F.L. thanks the Institut Universitaire de France as well as the EMBRC platform PIQs for image analysis. S. Sunagawa is supported by ETH Zürich and the Helmut Horten Foundation and by funding from the Swiss National Foundation (205321_184955). The authors declare that all data reported herein are fully and freely available from the date of publication, with no restrictions, and that all of the analyses, publications and ownership of data are free from legal entanglement or restriction by the various nations in whose waters the *Tara Oceans* expeditions conducted sampling. This article is contribution number 100 of *Tara Oceans*.
- Author contributions**
 S. Sunagawa and C.d.V. are the lead authors of the article and all other authors contributed to discussion of the content, writing and editing of the article.
- Competing interests**
 The authors declare no competing interests.
- Peer review information**
Nature Reviews Microbiology thanks David Hutchins, Maria Pachiadaki and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
- Publisher's note**
 Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
- Supplementary information**
 Supplementary information is available for this paper at <https://doi.org/10.1038/s41579-020-0364-5>.
- RELATED LINKS**
 Artist profiles: <https://oceans.taraexpeditions.org/en/m/artists/>
 Plankton chronicles: <http://planktonchronicles.org>
 Tara Ocean Foundation: <https://oceans.taraexpeditions.org/en>
 Tara Oceans Sample Registry: <https://doi.pangaea.de/10.1594/PANGAEA.875582>
 Tara Oceans Sequencing: <https://www.ebi.ac.uk/ena/data/view/PRJEB402>
 UniEuk: <http://uniek.org>

Tara Oceans Coordinators

Silvia G. Acinas², Marcel Babin^{7,20}, Peer Bork^{3,4,5}, Emmanuel Boss²¹, Chris Bowler^{6,7}, Guy Cochrane²², Colombe de Vargas^{7,19}, Michael Follows²³, Gabriel Gorsky^{7,9}, Nigel Grimsley^{7,24,25}, Lionel Guidi^{7,9}, Pascal Hingamp^{7,26}, Daniele Iudicone¹⁰, Olivier Jaillon^{7,18}, Stefanie Kandels^{3,7}, Lee Karp-Boss²¹, Eric Karsenti^{6,7,11}, Magali Lescot^{7,26}, Fabrice Not¹⁹, Hiroyuki Ogata¹², Stéphane Pesant^{1,5,14}, Nicole Poulton²⁷, Jeroen Raes^{28,29,30}, Christian Sardet^{7,9}, Mike Sieracki²⁷, Sabrina Speich^{31,32}, Lars Stemmann^{7,9}, Matthew B. Sullivan^{15,16,17}, Shinichi Sunagawa¹ and Patrick Wincker^{7,18}

²⁰Département de Biologie, Québec Océan and Takuviik Joint International Laboratory (UMI 3376), Université Laval (Canada)–CNRS (France), Université Laval, Quebec, QC, Canada. ²¹School of Marine Sciences, University of Maine, Orono, ME, USA. ²²European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. ²³Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. ²⁴CNRS UMR 7232, Biologie Intégrative des Organismes Marins, Banyuls-sur-Mer, France. ²⁵Sorbonne Universités Paris 06, OOB UPMC, Banyuls-sur-Mer, France. ²⁶Aix Marseille Université, Université de Toulon, CNRS, IRD, MIO UM 110, Marseille, France. ²⁷Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA. ²⁸Department of Microbiology and Immunology, Rega Institute, KU Leuven, Leuven, Belgium. ²⁹Center for the Biology of Disease, VIB KU Leuven, Leuven, Belgium. ³⁰Department of Applied Biological Sciences, Vrije Universiteit Brussel, Brussels, Belgium. ³¹Department of Geosciences, Laboratoire de Météorologie Dynamique, École Normale Supérieure, Paris, France. ³²Ocean Physics Laboratory, University of Western Brittany, Brest, France.