OCEAN PLANKTON

Eukaryotic plankton diversity in the sunlit ocean

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

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

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

A global metabarcoding approach

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

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

[1]CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France. [2]Sorbonne Universités, Université Pierre et Marie Curie (UPMC) Paris 06, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France. [3]Department of Ecology, University of Kaiserslautern, Erwin-Schroedinger Street, 67663 Kaiserslautern, Germany. [4]Department of Marine Biology and Oceanography, Institute of Marine Science (ICM)–Consejo Superior de Investigaciones Científicas (CSIC), Passeig Marítim de la Barceloneta 37-49, Barcelona E08003, Spain. [5]Laboratory of Soil Biology, University of Neuchâtel, Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland. [6]CNRS, FR2424, Roscoff Culture Collection, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France. [7]Sorbonne Universités, UPMC Paris 06, FR 2424, Roscoff Culture Collection, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France. [8]Ecole Normale Supérieure, Institut de Biologie de l'ENS (IBENS), and Inserm U1024, and CNRS UMR 8197, Paris, F-75005 France. [9]Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Institut de Génomique, GENOSCOPE, 2 rue Gaston Crémieux, 91000 Evry, France. [10]CNRS FR3631, Institut de Biologie Paris-Seine, F-75005, Paris, France. [11]Sorbonne Universités, UPMC Paris 06, Institut de Biologie Paris-Seine, F-75005, Paris, France. [12]Department of Microbiology and Immunology, Rega Institute, KU Leuven, Herestraat 49, 3000 Leuven, Belgium. [13]Center for the Biology of Disease, VIB, Herestraat 49, 3000 Leuven, Belgium. [14]Department of Applied Biological Sciences, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. [15]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Branišovská 31, 37005 České Budějovice, Czech Republic. [16]Faculty of Science, University of South Bohemia, Branišovská 31, 37005 České Budějovice, Czech Republic. [17]CNRS, UMR 7093, Laboratoire d'Océanographie de Villefranche-sur-Mer (LOV), Observatoire Océanologique, F-06230, Villefranche-sur-Mer, France. [18]Sorbonne Universités, UPMC Paris 06, UMR 7093, LOV, Observatoire Océanologique, F-06230, Villefranche-sur-Mer, France. [19]CNRS, UMR 8030, CP5706, Evry, France. [20]Université d'Evry, UMR 8030, CP5706, Evry, France. [21]Canadian Institute for Advanced Research, 180 Dundas Street West, Suite 1400, Toronto, Ontario M5G 1Z8, Canada. [22]MARUM, Center for Marine Environmental Sciences, University of Bremen, 28359 Bremen, Germany. [23]Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy. [24]Ifremer, Centre de Brest, DYNECO/Pelagos CS 10070, 29280 Plouzané, France. [25]Structural and Computational Biology, European Molecular Biology Laboratory (EMBL), Meyerhofstraße 1, 69117 Heidelberg, Germany. [26]Directors' Research, EMBL, Meyerhofstraße 1, 69117 Heidelberg, Germany. [27]Max-Delbrück-Centre for Molecular Medicine, 13092 Berlin, Germany. [28]CNRS UMR 7232, Biologie Intégrative des Organismes Marins (BIOM), Avenue du Fontaulé, 66650 Banyuls-sur-Mer, France. [29]Sorbonne Universités Paris 06, Observatoire Océanologique de Banyuls (OOB) UPMC, Avenue du Fontaulé, 66650 Banyuls-sur-Mer, France. [30]Aix Marseille Université, CNRS IGS UMR 7256, 13288 Marseille, France. [31]Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan. [32]PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, Bremen, Germany. [33]Bigelow Laboratory for Ocean Sciences, East Boothbay, ME 04544, USA. [34]National Science Foundation, Arlington, VA 22230, USA. [35]Department of Geosciences, Laboratoire de Météorologie Dynamique (LMD), Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05, France. [36]Laboratoire de Physique des Océans, Université de Bretagne Occidentale (UBO)–Institut Universitaire Européen de la Mer (IUEM), Place Copernic, 29820 Plouzané, France.

***Corresponding author. E-mail: vargas@sb-roscoff.fr (C.d.V.); pwincker@genoscope.cns.fr (P.W.); karsenti@embl.de (E.K.)**

†These authors contributed equally to this work. ‡*Tara* Oceans Coordinators and affiliations appear at the end of this paper.