

Determinants of community structure in the global plankton interactome

Gipsi Lima-Mendez,^{1,2,3*} Karoline Faust,^{1,2,3*} Nicolas Henry,^{4,5*} Johan Decelle,^{4,5} Sébastien Colin,^{4,5,6} Fabrizio Carcillo,^{1,2,3,7} Samuel Chaffron,^{1,2,3} J. Cesar Ignacio-Espinosa,^{8†} Simon Roux,^{8†} Flora Vincent,^{2,6} Lucie Bittner,^{4,5,6,9} Youssef Darzi,^{2,3} Jun Wang,^{1,2} Stéphane Audic,^{4,5} Léo Berline,^{10,11} Gianluca Bontempi,⁷ Ana M. Cabello,¹² Laurent Coppola,^{10,11} Francisco M. Cornejo-Castillo,¹² Francesco d'Ovidio,¹³ Luc De Meester,¹⁴ Isabel Ferrera,¹² Marie-José Garet-Delmas,^{4,5} Lionel Guidi,^{10,11} Elena Lara,¹² Stéphane Pesant,^{15,16} Marta Royo-Llonch,¹² Guillem Salazar,¹² Pablo Sánchez,¹² Marta Sebastian,¹² Caroline Souffreau,¹⁴ Céline Dimier,^{4,5,6} Marc Picheral,^{10,11} Sarah Searson,^{10,11} Stefanie Kandels-Lewis,^{17,18} Tara Oceans coordinators[†], Gabriel Gorsky,^{10,11} Fabrice Not,^{4,5} Hiroyuki Ogata,¹⁹ Sabrina Speich,^{20,21} Lars Stemann,^{10,11} Jean Weissenbach,^{22,23,24} Patrick Wincker,^{22,23,24} Silvia G. Acinas,¹² Shinichi Sunagawa,¹⁷ Peer Bork,^{17,25} Matthew B. Sullivan,^{8†} Eric Karsenti,^{6,18§} Chris Bowler,^{6§} Colman de Vargas,^{4,5§} Jeroen Raes^{1,2,3§}

Species interaction networks are shaped by abiotic and biotic factors. Here, as part of the *Tara Oceans* project, we studied the photic zone interactome using environmental factors and organismal abundance profiles and found that environmental factors are incomplete predictors of community structure. We found associations across plankton functional types and phylogenetic groups to be nonrandomly distributed on the network and driven by both local and global patterns. We identified interactions among grazers, primary producers, viruses, and (mainly parasitic) symbionts and validated network-generated hypotheses using microscopy to confirm symbiotic relationships. We have thus provided a resource to support further research on ocean food webs and integrating biological components into ocean models.

The structure of oceanic ecosystems results from the complex interplay between resident organisms and their environment. In the world's largest ecosystem, oceanic plankton (composed of viruses, prokaryotes, microbial eukaryotes, phytoplankton, and zooplankton) form trophic and symbiotic interaction networks (1–4) that are influenced by environmental conditions. Ecosystem structure and composition are governed by abiotic as well as biotic factors. The former include environmental conditions and nutrient availability (5), whereas the latter include grazing, pathogenicity, and parasitism (6, 7). Historically, abiotic factors have been considered to have a stronger effect, but recently, appreciation for biotic factors is growing (8, 9). We sought to develop a quantitative understanding of biotic and abiotic interactions in natural systems in which the organisms are taxonomically and trophically diverse (10). We used sequencing technologies to profile communities across trophic levels, organismal sizes, and geographic ranges and to predict organismal interactions across biomes based on co-occurrence patterns (11). Previous efforts addressing these issues have provided insights on the structure (12, 13) and dynamics of microbial communities (14–16).

We analyzed data from 313 plankton samples the *Tara Oceans* expedition (17) derived from seven size-fractions covering collectively 68 stations at two depths across eight oceanic provinces (table S1). The plankton samples spanned sizes

that include organisms from viruses to small metazoans. We derived viral, prokaryotic, and eukaryotic abundance profiles from clusters of metagenomic contigs, Illumina-sequenced metagenomes (*mitags*), and 18S ribosomal DNA (rDNA) V9 sequences, respectively (table S1) (10, 18, 19) and collected environmental data from on-site and satellite measurements (17, 20, 21). We used network inference methods and machine-learning techniques so as to disentangle biotic and abiotic signals shaping ocean plankton communities and to construct an interactome that described the network of interactions among photic zone plankton groups. We used the interactome to focus on specific relationships, which we validated through microscopic analysis of symbiont pairs and in silico analysis of phage-host pairings.

Evaluating the effect of abiotic and biotic factors on community structure

We first reassessed the effects of environment and geography on community structure. Using variation partitioning (22), we found that on average, the percentage of variation in community composition explained by environment alone was 18%, by environment combined with geography 13%, and by geography alone only 3% (23, 24). In addition, we built random forest-based models (25) in order to predict abundance profiles of the Operational Taxonomic Units (OTU) using (i) OTUs alone, (ii) environmental variables alone, and (iii) OTUs and environmental variables combined and tested for each OTU whether one of

the three approaches outcompeted the other. These analyses revealed that 95% of the OTU-only models are more accurate in predicting OTU abundances than environmental variable models, and that combined models were no better than the OTU-only models (26, 27). This suggests that abiotic factors have a more limited effect on community structure than previously assumed (8).

To study the role of biotic interactions, we developed a method with which to identify robust species associations in the context of environmental conditions. Twenty-three taxon-taxon and taxon-environment co-occurrence networks were constructed based on 9292 taxa, representing the combinations of two depths, seven organismal

¹Department of Microbiology and Immunology, Rega Institute KU Leuven, Herestraat 49, 3000 Leuven, Belgium. ²VIB Center for the Biology of Disease, VIB, Herestraat 49, 3000 Leuven, Belgium. ³Department of Applied Biological Sciences (DBIT), Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. ⁴Station Biologique de Roscoff, CNRS, UMR 7144, Place Georges Teissier, 29680 Roscoff, France. ⁵Sorbonne Universités, Université Pierre et Marie Curie (UPMC) Université Paris 06, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France. ⁶Ecole Normale Supérieure, Institut de Biologie de l'ENS (IBENS), Inserm U1024, CNRS UMR 8197, Paris, F-75005 France. ⁷Interuniversity Institute of Bioinformatics in Brussels (IB)², ULB Machine Learning Group, Computer Science Department, Université Libre de Bruxelles (ULB), Brussels, Belgium. ⁸Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, 85721, USA. ⁹Institut de Biologie Paris-Seine, CNRS FR3631, F-75005, Paris, France. ¹⁰CNRS, UMR 7093, Laboratoire d'Océanographie de Villefranche (LOV), Observatoire Océanologique, F-06230 Villefranche-sur-mer, France. ¹¹Sorbonne Universités, UPMC Paris 06, UMR 7093, Laboratoire d'Océanographie de Villefranche (LOV), Observatoire Océanologique, F-06230 Villefranche-sur-mer, France. ¹²Department of Marine Biology and Oceanography, Institute of Marine Sciences (ICM)–Consejo Superior de Investigaciones Científicas (CSIC), Pg. Marítim de la Barceloneta, 37-49, Barcelona E08003, Spain. ¹³Sorbonne Universités, UPMC, Université Paris 06, CNRS–Institut pour la Recherche et le Développement–Muséum National d'Histoire Naturelle, Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques (LOCEAN) Laboratory, 4 Place Jussieu, 75005, Paris, France. ¹⁴KU Leuven, Laboratory of Aquatic Ecology, Evolution and Conservation, Charles Deberiotstraat 32, 3000 Leuven. ¹⁵PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, Hochschule 18, 28359 Bremen, Germany. ¹⁶MARUM, Center for Marine Environmental Sciences, University of Bremen, Hochschule 18, 28359 Bremen, Germany. ¹⁷Structural and Computational Biology, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. ¹⁸Directors' Research, European Molecular Biology Laboratory, Heidelberg, Germany. ¹⁹Institute for Chemical Research, Kyoto University, Gokasho, Uji, 611-0011 Kyoto, Japan. ²⁰Department of Geosciences, Laboratoire de Météorologie Dynamique (LMD), Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05, France. ²¹Laboratoire de Physique des Océans, Université de Bretagne Occidentale (UBO)–Institut Universitaire Européen de la Mer (IUEM), Palce Copernic, 29820 Plouzané, France. ²²Commissariat à l'Énergie Atomique (CEA), Genoscope, 2 rue Gaston Crémieux, 91000 Evry, France. ²³CNRS, UMR 8030, 2 rue Gaston Crémieux, 91000 Evry, France. ²⁴Université d'Evry, UMR 8030, CP5706 Evry, France. ²⁵Max-Delbrück-Centre for Molecular Medicine, 13092 Berlin, Germany. *These authors contributed equally to this work. †Present address: Department of Microbiology, Ohio State University, Columbus, OH 43210, USA. ‡Tara Oceans coordinators and affiliations are listed at the end of this manuscript. §Corresponding author. E-mail: jeroen.raes@vib-kuleuven.be (J.R.); vargas@sb-roscoff.fr (C.d.V.); cbowler@biologie.ens.fr (C.B.); karsenti@embl.de (E.K.)