OCEAN PLANKTON

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 Stemmann,[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]§ Colomban 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 ($_{mi}$tags), and 18*S* 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

[1]Department of Microbiology and Immunology, Rega Institute KU Leuven, Herestraat 49, 3000 Leuven, Belgium. [2]VIB Center for the Biology of Disease, VIB, Herestraat 49, 3000 Leuven, Belgium. [3]Department of Applied Biological Sciences (DBIT), Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. [4]Station Biologique de Roscoff, CNRS, UMR 7144, Place Georges Teissier, 29680 Roscoff, France. [5]Sorbonne Universités, Université Pierre et Marie Curie (UPMC) Université Paris 06, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France. [6]Ecole Normale Supérieure, Institut de Biologie de l'ENS (IBENS), Inserm U1024, CNRS UMR 8197, Paris, F-75005 France. [7]Interuniversity Institute of Bioinformatics in Brussels (IB)^2, ULB Machine Learning Group, Computer Science Department, Université Libre de Bruxelles (ULB), Brussels, Belgium. [8]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, 85721, USA. [9]Institut de Biologie Paris-Seine, CNRS FR3631, F-75005, Paris, France. [10]CNRS, UMR 7093, Laboratoire d'Océanographie de Villefranche (LOV), Observatoire Océanologique, F-06230 Villefranche-sur-mer, France. [11]Sorbonne Universités, UPMC Paris 06, UMR 7093, Laboratoire d'Océanographie de Villefranche (LOV), Observatoire Océanologique, F-06230 Villefranche-sur-mer, France. [12]Department 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. [13]Sorbonne 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. [14]KU Leuven, Laboratory of Aquatic Ecology, Evolution and Conservation, Charles Deberiotstraat 32, 3000 Leuven. [15]PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, Hochschulring 18, 28359 Bremen, Germany. [16]MARUM, Center for Marine Environmental Sciences, University of Bremen, Hochschulring 18, 28359 Bremen, Germany. [17]Structural and Computational Biology, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. [18]Directors' Research, European Molecular Biology Laboratory, Heidelberg, Germany. [19]Institute for Chemical Research, Kyoto University, Gokasho, Uji, 611-0011 Kyoto, Japan. [20]Department of Geosciences, Laboratoire de Météorologie Dynamique (LMD), Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05, France. [21]Laboratoire de Physique des Océan, Université de Bretagne Occidentale (UBO)–Institut Universaire Européen de la Mer (IUEM), Palce Copernic, 29820 Polouzané, France. [22]Commissariat à l'Énergie Atomique (CEA), Genoscope, 2 rue Gaston Crémieux, 91000 Evry, France. [23]CNRS, UMR 8030, 2 rue Gaston Crémieux, 91000 Evry, France. [24]Université d'Evry, UMR 8030, CP5706 Evry, France. [25]Max-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.)**

size ranges, and four organismal domains (Bacteria, Archaea, Eukarya, and viruses) (28). To reduce noise and thus false-positive predictions, we restricted our analysis to taxa present in at least 20% of the samples and used conservative statistical cutoffs. We merged the individual networks into a global network, which features a total of 127,995 distinct edges, of which 92,633 are taxon-taxon edges and 35,362 are taxon-environment edges (Table 1). Node degree does not depend on the abundance of the node (28). As such, this network represents a resource with which to examine species associations in the global oceans (28–31).

Next, we assessed how many of the taxon links represented "niche effects" driven by geography or environment (such as when taxa respond similarly to a common environmental condition). We examined motifs consisting of two correlated taxa that also correlate with at least one common environmental parameter ("environmental triplets" to identify associations that were driven by environment) using three approaches [interaction information, sign pattern analysis, and network deconvolution (32)]. We identified 29,912 taxon-taxon-environment associations (32.3% of total). Among environmental factors, we found that PO_4, temperature, NO_2, and mixed-layer depth were frequent drivers of network connections (Fig. 1A). Although the three methodologies pinpoint indirect associations, only interaction information directly identifies synergistic effects in these biotic-abiotic triplets. Exploiting this property, we disentangled the 29,912 environment-affected associations into 11,043 edges driven solely by abiotic factors (excluded from the network for the remainder of the study) (31, 33) and 18,869 edges whose dependencies result from biotic-abiotic synergistic effects. Thus, we find that a minority of associations can be explained by an environmental factor.

Evaluation of predicted interactions

Co-occurrence techniques have heretofore mainly been applied to bacteria. We detected eukaryotic interactions on the basis of analysis of sequences at the V9 hypervariable region of the 18*S* ribosomal RNA (rRNA) gene. We built a literature-curated collection (34) of 574 known symbiotic interactions (including both parasitism and mutualism) in marine eukaryotic plankton (30, 35). From 43 genus-level interactions represented by OTUs in the abundance preprocessed input matrices, we found 42% (18 genus pairs; 47% when limiting to parasitic interactions) represented in our reference list. The probability of having found each of these interactions by chance alone was <0.01 (Fisher exact test, average $P = 4^{-3}$, median $P = 5e^{-7}$). On the basis of this sensitivity and a false discovery rate averaging to 9% (computed from null models), we estimate the number of interactions among eukaryotes present in our filtered input matrices to be between 53,000 and 139,000. Most of the false-negative interactions were due to the strict filtering rules we used to avoid false positives; this hampers detection when, for example, interactions are facultative or when interaction partners may vary among closely related groups depending on oceanic region (4). False positives could represent indirect interactions between species (bystander effects) or environmental effects caused by factors not captured in this study (36, 37).

Biotic interactions within and across kingdoms

The integrated network contained 81,590 predicted biotic interactions (30) that were nonrandomly distributed within and between size fractions (Fig. 1, B and C) (38). Positive associations outnumbered mutual exclusions (72% versus 28%), and we observed a nonrandom edge distribution with regard to phylogeny (Fig. 2A), with most associations derived from syndiniales and other dinoflagellates (examples are shown in

Table 1. Properties of the merged taxon network. The positive subset of the network was clustered with the leading eigen vector algorithm (91).

Nodes	Edges	Positive edges (%)	Negative edges	Average clustering coefficient	Average path length	Diameter	Average betweenness	Modularity of positive network	Number of modules in positive network
9169	92,633	68,856 (74.33)	23,777	0.229	3.43	12	11024	0.51	51

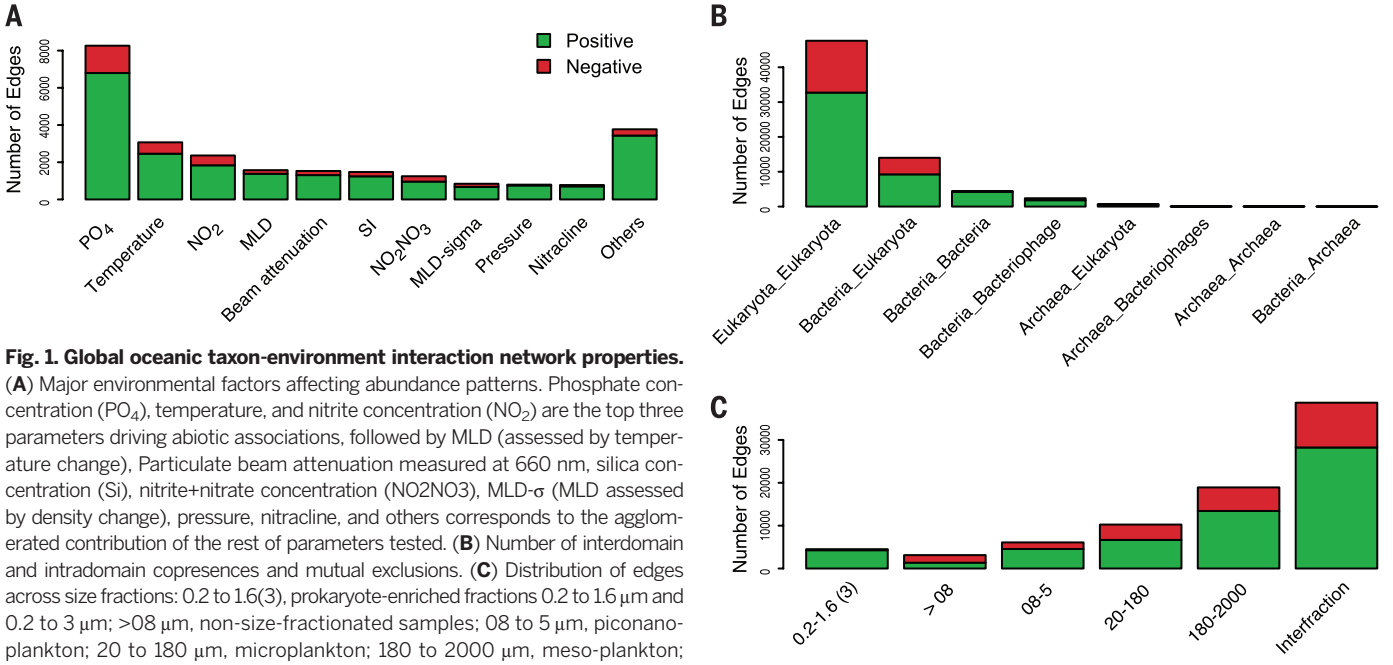

Fig. 1. Global oceanic taxon-environment interaction network properties. (**A**) Major environmental factors affecting abundance patterns. Phosphate concentration (PO_4), temperature, and nitrite concentration (NO_2) are the top three parameters driving abiotic associations, followed by MLD (assessed by temperature change), Particulate beam attenuation measured at 660 nm, silica concentration (Si), nitrite+nitrate concentration (NO2NO3), MLD-σ (MLD assessed by density change), pressure, nitracline, and others corresponds to the agglomerated contribution of the rest of parameters tested. (**B**) Number of interdomain and intradomain copresences and mutual exclusions. (**C**) Distribution of edges across size fractions: 0.2 to 1.6(3), prokaryote-enriched fractions 0.2 to 1.6 μm and 0.2 to 3 μm; >08 μm, non-size-fractionated samples; 08 to 5 μm, piconanoplankton; 20 to 180 μm, microplankton; 180 to 2000 μm, meso-plankton; interfrac, includes interfraction networks 08 to 5 μm versus 20 to 180 μm, 08 to 5 μm versus 180 to 2000 μm, 20 to 180 μm versus 180 to 2000 μm, and 0.2 to 1.6(3) μm versus ≤ 0.2 μm (virus-enriched fraction).

Fig. 3A), and exclusions involving arthropods. Certain combinations of phylogenetic groups are overrepresented (*39*). For instance, we found a clade of syndiniales [the MALV-II Clade 1 belonging to *Amoebophrya* (*3*)] enriched in positive associations with tintinnids ($P = 2^{-4}$), which are among the most abundant ciliates in marine plankton (*40*). The tintinnid *Xystonella lohmani* was described in 1964 to be infected by *Amoebophrya tintinnis* (*41*), and tintinnids can feed on *Amoebophrya* free-living stages (*42*). Other found host-parasite associations included the copepod parasites *Blastodinium*, *Ellobiopsis*, and *Vampyrophrya* (*41*, *43*–*45*).

On the other hand, *Maxillopoda*, *Bacillariophyceae*, and collodarians, three groups of relatively large sized organisms whose biomass can dominate planktonic ecosystems, are rich in negative associations among them (*33*). Collodarians and copepods are abundant in, respectively, the oligotrophic tropical and eutrophic and mesotrophic temperate systems (*10*, *46*). The decoupling of phyto- and zooplankton in open oceans by diatoms anticorrelating to copepods (*47*, *48*) is attributed to growth rate differences and to the diatom production of compounds harmful to their grazers (*49*). The combination of these

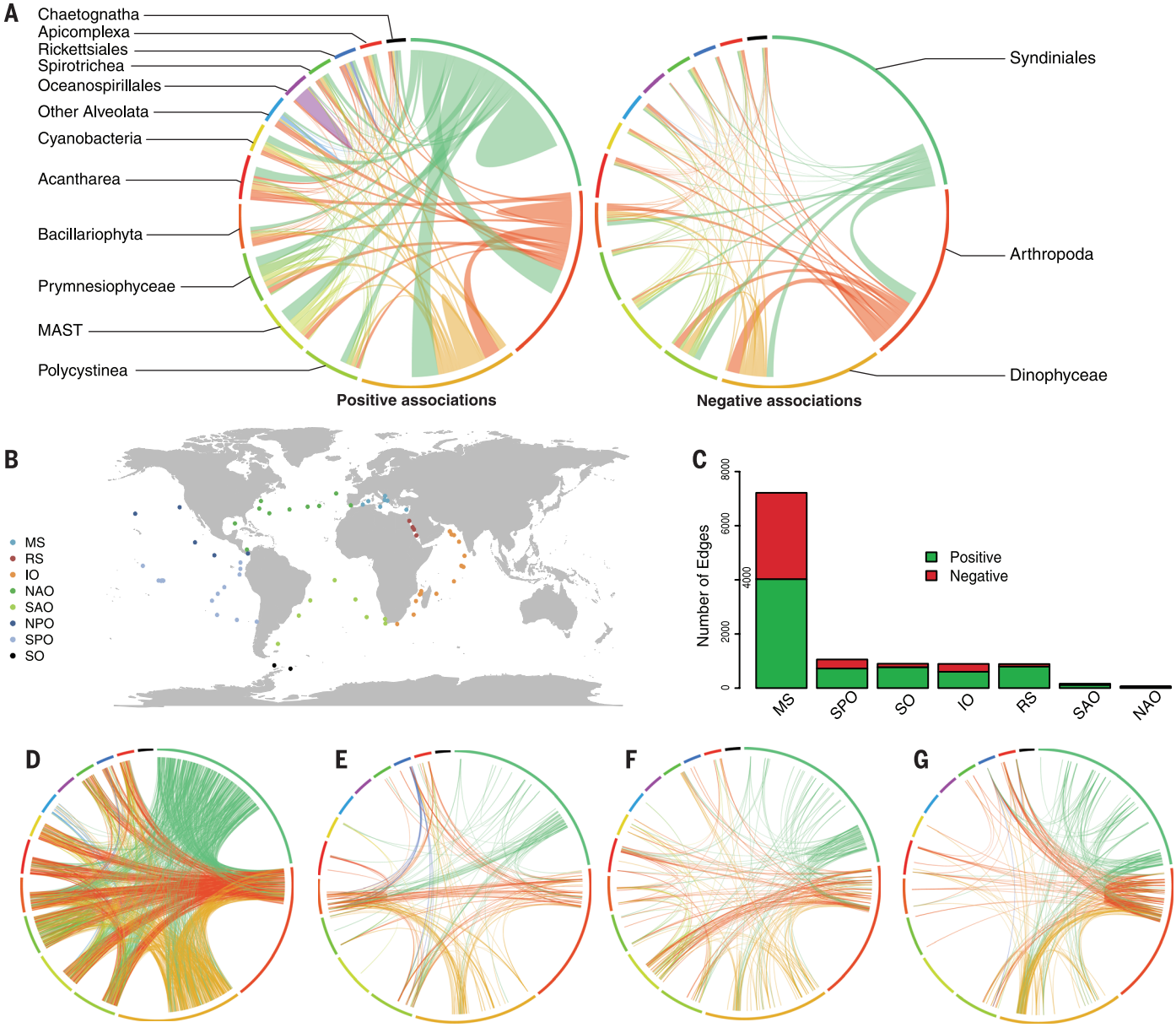

Fig. 2. Taxonomic and geographic patterns within the co-occurrence network. (**A**) Top 15 interacting taxon groups depicted as colored segments in a CIRCOS plot, in which ribbons connecting two segments indicate co-presence and exclusion links, on the left and right, respectively. Size of the ribbon is proportional to the number of links (copresences and exclusions) between the OTUs assigned to the respective segments, and color is segment (of the two involved) with the more total links. Links are dominated by the obligate parasites syndiniales and by Arthropoda and Dinophyceae. (**B**) *Tara* Oceans sampling stations grouped by oceanic provinces. (**C**) Frequency of local co-occurrence patterns across the oceanic provinces, showing that most local patterns are located in MS. (**D** to **G**) Taxonomic patterns of co-occurrences across MS (D), SPO (E), IO (F), and RS (G). Edges are represented as ribbons between barcodes grouped into their taxonomic order as in (A). Links sharing the same segment are affiliated to the same taxon (Order), showing that the connectivity patterns across taxa are conserved at high taxonomic ranks. The local specificity of interactions at higher resolution (OTUs) is apparent by thin ribbons (edge resolution), with different starts, and end positions (different OTUs) within the shared (taxon) segment, section color, and ordering correspond to those in (A). SO-specific associations are mainly driven by bacterial interactions (*53*).

effects could lie at the basis of this observation, which contrasts with other free-living autotrophs represented in the network (cyanobacteria and prymnesiophytes), which display primarily positive associations (Fig. 2A).

Cross-kingdom associations between Bacteria and Archaea were limited to 24 mutual exclusions. Within Archaea, Thermoplasmatales (Marine Group II) co-occur with several phytoplankton clades. Links between Bacteria and protists recovered five out of eight recently discovered interactions from protist single-cell sequencing (*50*). Associations between Diatoms and Flavobacteria agreed with their described symbioses (*51*). We also observed co-ocurrence of uncultured dinoflagellates with members of Rhodobacterales (*Ruegeria*), which is in agreement with a symbiosis between *Ruegeria* sp. TM1040 and *Pfiesteria piscicida* around the ability of *Ruegeria* to metabolize dinoflaggelate-produced dimethylsulfoniopropionate (*52*).

Global versus local associations

We further investigated whether our network was driven by global trends or is defined by local signals. To this aim, we divided our set of samples into seven main regions—Mediterranean

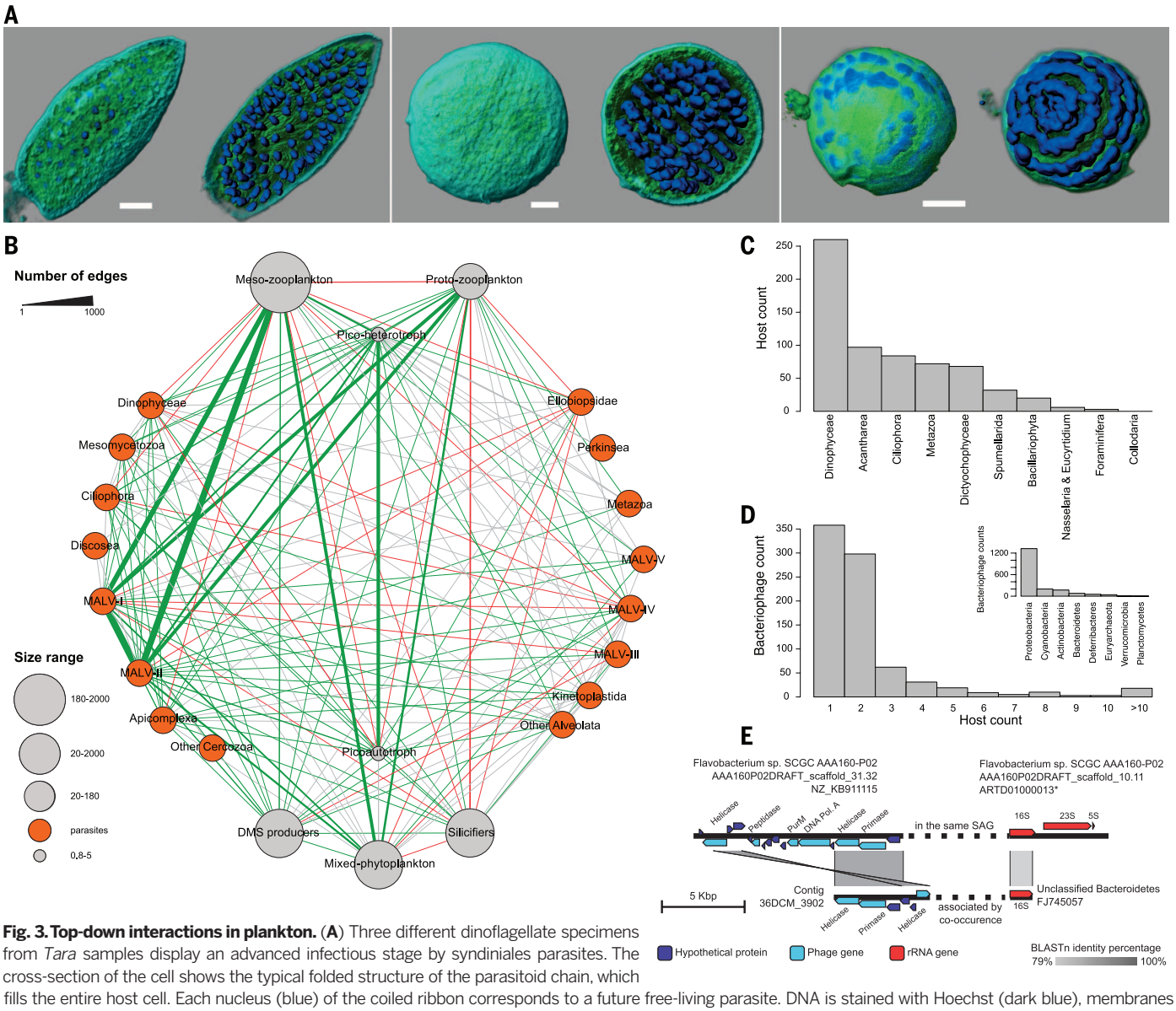

Fig. 3. Top-down interactions in plankton. (**A**) Three different dinoflagellate specimens from *Tara* samples display an advanced infectious stage by syndiniales parasites. The cross-section of the cell shows the typical folded structure of the parasitoid chain, which fills the entire host cell. Each nucleus (blue) of the coiled ribbon corresponds to a future free-living parasite. DNA is stained with Hoechst (dark blue), membranes are stained with DiOC6 (green), and specimen surface is light blue. Scale bar, 5 μm. (**B**) Subnetwork of metanodes that encapsulate barcodes affiliated to parasites or PFTs. The PFTs mapped onto the network are: phytoplankton DMS producers, mixed phytoplankton, phytoplankton silicifiers, pico-eukaryotic heterotrophs, proto-zooplankton and meso-zooplankton. Edge width reflects the number of edges in the taxon graph between the corresponding metanodes. Over-represented links (multiple-test corrected $P < 0.05$) are colored in green if they represent copresences and in red if they represent exclusions; gray means non-overrepresented combinations. When both copresences and exclusions were significant, the edge is shown as copresence. (**C**) Parasite connections within micro- and zooplankton groups. (**D**) Number of hosts per phage. (Inset) Phage associations to bacterial (target) phyla. (**E**) Putative Bacteroidetes viruses detected with co-occurence and detection in a single-cell genome (SAG). On the left are viral sequences from a Flavobacterium SAG (top) and *Tara* Oceans virome (bottom), displaying an average of 89% nucleotide identity. On the right is the correspondence between the ribosomal genes detected in the same SAG (top) and the 16*S* sequence associated to the *Tara* Oceans contig based on co-occurence (79% nucleotide identity). For clarity, a subset of contig ARTD0100013 only (from 10,000 to 16,000 nucleotides) is displayed. This sequence was also reverse-complemented. PurM, phosphoribosylaminoimidazole synthetase; DNA Pol. A, DNA polymerase A.

Sea (MS), Red Sea (RS), Indian Ocean (IO), South Atlantic (SAO), Southern Ocean (SO), South Pacific Ocean (SPO) and North Atlantic Ocean (NAO)—and assessed the "locality" of associations by comparing the score with or without that region. We found that association patterns were mostly driven by global trends because only 14% of edges were identified as local (Fig. 2, B and C). Approximately two thirds of local associations occur in MS (7215), followed by SPO (1058), whereas the rest are contributed by SO (901), IO (894), RS (889), SAO (163), and NAO (60) (Fig. 2, C to G). MS was the region with most sampling sites, which allowed us to recover more local patterns. Nevertheless, Fig. 2, C to G, shows that although the same major groups (order level) interact in both the global and local networks, each local site has its own specific interaction profile ($P < 1^{-8}$) (*33*, *39*, *53*).

Parasite impact on plankton functional types

Parasitic interactions are the most abundant pattern present in the network, which is also eminent by repeated microscopic observation of parasitic interactions from the *Tara* samples (Fig. 3A). We focused on predicted parasitic interactions and assessed their potential impact on biogeochemical processes by exploring a functional subnetwork (21,572 edges) of known and previously unidentified plankton parasites (*10*) together with classical "plankton functional types" (PFTs) (*54*). PFTs group taxa by trophic strategy (for example, autotrophs versus heterotrophs) and role in ocean biogeochemistry (Fig. 3A) (*55*). The relationship between the different PFTs (network density of 0.65) highlights strong dependencies between phytoplankton and grazers. We found that all PFTs are associated with parasites, but not always to the same extent. Most links involve syndiniales MALV-I and MALV-II clades associated to zooplankton and, to a lesser extent, to microphytoplankton (excluding diatoms). This emphasizes the role of alveolate parasitoids as top-down effectors of zooplankton and microphytoplankton population structure and functioning (*3*), although the latter group is also affected by grazing (*1*). The meso-planktonic networks contain known syndiniales targets (Dinophyceae, Ciliophora, Acantharia, and Metazoa) (Fig. 3B) (*56*). In large size fractions, we found interactions between known parasites and groups of organisms that in theory are too small to be their hosts (*57*); 32% of these associations involved the abundant and diverse marine stramenopiles (MASTs) and diplonemids (other Discoba and Diplonema) (*10*). Ecophysiology studies (*58*, *59*) suggest a parasitic role for these lineages. The association of these groups with other parasites would be explained by putative co-infection of the same hosts. Contrasting with the above observations, we found phytoplankton silicifiers (diatoms) displaying a variety of mutual exclusions. One possible interpretation of this is that diatom silicate exoskeletons (*60*) and toxic compound production (*49*) could act as efficient barriers against top-down pressures (*61*).

Phage-microbe associations

We investigated phage-microbe interactions, another major top-down process affecting global bacterial/archaeal community structure (*7*). Here, surface (SRF) and deep chlorophyll maximum (DCM) virus-bacteria networks revealed 1869 positive associations between viral populations and 7 of the 54 known bacterial phyla (specifically, Proteobacteria, Cyanobacteria, Actinobacteria, Bacteroidetes, Deferribacteres, Verrucomicrobia, and Planctomycetes), and one archaeal phylum (Euryarchaeota). These eight phyla represent most of abundant bacterial/archaeal groups across 37 investigated samples (Fig. 3D), suggesting that the networks are detecting abundant virus-host interactions. Additionally, these interactions include phyla of microbes lacking viral genomes in RefSeq databases including Verrucomicrobia, and nonextremophile Euryarchaeota, hinting at genomic sequences for understudied viral taxa (Fig. 3E) (*39*, *62*, *63*). Among the phage populations in the network, we found eight corresponding to phage sequences available in GenBank (>50% of genes with a >50% amino acid identity match). In all eight cases, the predicted host from the network corresponded to the annotated host family in the GenBank record, which is significantly higher than expected by chance ($P = 0.001$) (*62*).

Next, we evaluated viral host range, which is fundamental for predictive modeling and thus far largely limited to observations of cultured virus-host systems that insufficiently map complex community interactions (*64*). Our virus-host interaction data suggest that viruses are very host-specific: ~43% of the phage populations interact with only a single host OTU, and the remaining 57% interact with only a few, often closely related OTUs (Fig. 3D). These networks are modular at large scales (*65*), suggesting that viruses are host range–limited across large sections of host space. Nestedness analysis showed inconsistent results across algorithms.

Microscopic validation of predicted interactions

Our data predicted a photosymbiotic interaction between an acoel flatworm (*Symsagittifera* sp.) and a green microalga (*Tetraselmis* sp.). We validated this by means of laser scanning confocal microscopy (LSCM), three-dimensional (3D) reconstruction, and reverse molecular identification on flatworm specimens isolated from *Tara* Oceans preserved morphological samples. We observed microalgal cells (5 to 10 μm in diameter) within each of the 15 isolated acoel specimens (Fig. 4) (*66*). The 18*S* sequence from several sorted holobionts matched the metabarcode pair identified in the co-occurrence global network. Thus, molecular ecology, bioinformatics, and microscopic analysis can enable the discovery of marine symbioses.

Conclusions

The global ocean interactome can be used to predict the dynamics and structure of ocean ecosystems. The interactome reported here spans all three organismal domains and viruses. The analyses presented emphasize the role of top-down biotic interactions in the epipelagic zone. This data will inform future research to understand how symbionts, pathogens, predators, and parasites interact with their target organisms and will ultimately help elucidate the structure of the global food webs that drive nutrient and energy flow in the ocean.

Methods

Sampling

The sampling strategy used in the *Tara* Oceans expedition is described in (*67*), and samples used in the present study are listed in table S1 and http://doi.pangaea.de/10.1594/PANGAEA.840721. The *Tara* Oceans nucleotide sequences are available at the European Nucleotide Archive (ENA) under projects PRJEB402 and PRJEB6610.

Physical and environmental measurements

Physical and environmental measurements were carried out with a vertical profile sampling system (CTD-rosette) and data collected from Niskin bottles. We measured temperature, salinity, chlorophyll, CDOM fluorescence (fluorescence of the colored dissolved organic matter), particles abundance, nitrate concentration, and particle size distribution (using an underwater vision profiler). In addition, mean mixed-layer depth (MLD), maximum fluorescence, vertical maximum of the Brünt-Väisälä Frequency N ($s - 1$), vertical range of dissolved oxygen, and depth of nitracline were determined. Satellite altimetry provided the Okubo-Weiss parameter, Lyapunov exponent, mesoscale eddie retention, and sea-surface temperature (SST) gradients at eddie fronts (*19*). Data are available at http://www.pangaea.de (http://doi.pangaea.de/10.1594/PANGAEA.840718).

Abundance table construction

Prokaryotic 16*S* rDNA metagenomic reads were identified, annotated, and quantified from $_{mi}$tags) as described in (*68*) by using the SILVA v.115 database (*19*, *69*, *70*). The abundance table was normalized by using the summed read count per sample (*19*, *71*). Quality-checked V9 rDNA metabarcodes were clustered into swarms as in (*10*, *72*) and annotated by using the V9 PR2 database (*73*). PR2 barcodes were associated to fundamental trophic modes (auto- or heterotrophy) and symbiotic interactions (parasitism and mutualism) according to literature (Taxonomic and trophic mode annotations are available at http://doi.pangaea.de/10.1594/PANGAEA.843018 and http://doi.pangaea.de/10.1594/PANGAEA.843022). Swarm abundance and normalization was performed as in (*10*, *72*). Bacteriophage metagenomes were obtained from the < 0.2-μm fractions for 48 samples, and contigs were annotated and quantified as in (*18*). The abundance matrix was normalized by means of total sample read count and contig length.

In all cases, only OTUs with relative abundance $> 1^{-8}$ and detected in at least 20% of samples were retained. Because sample number in the input tables ranged from 17 to 63, prevalence thresholds varied (from 22 to 40%). The sum of all filtered OTU relative abundances was kept in the tables to preserve proportions. Abundance tables

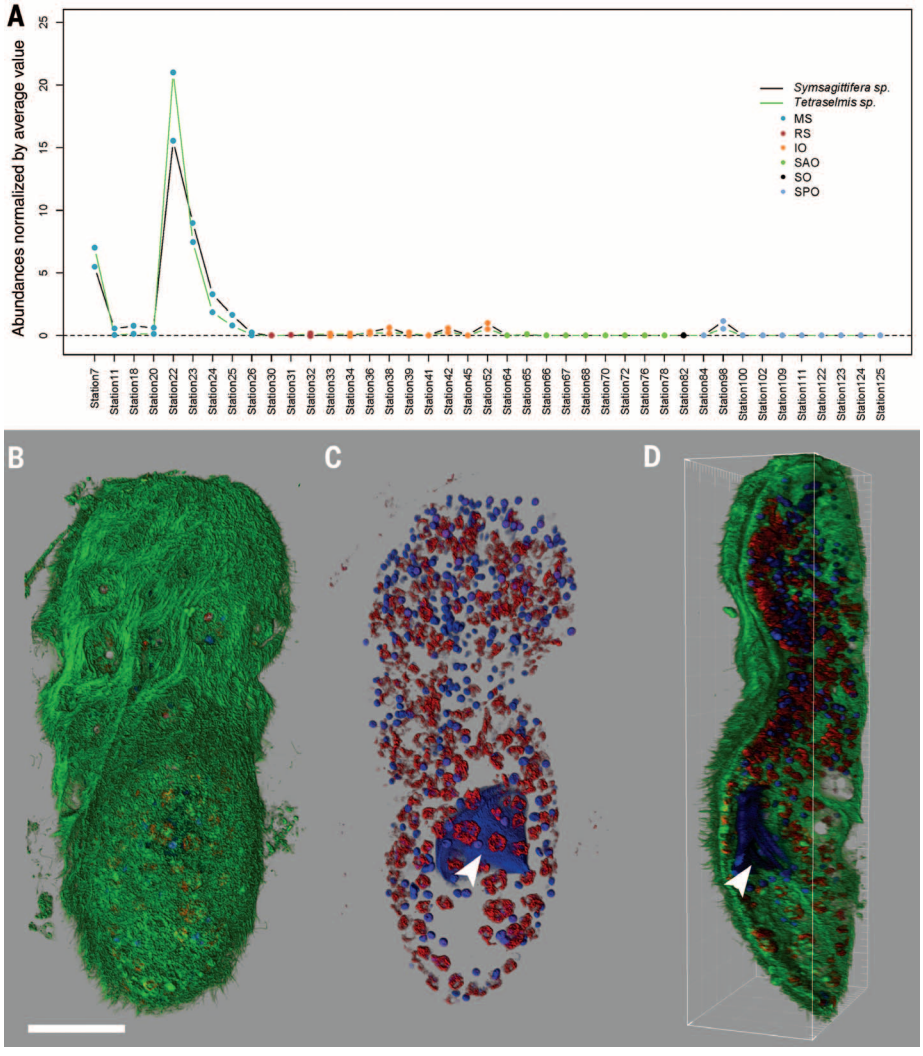

Fig. 4. Experimental validation of network-predicted interaction (photosymbiosis). Guided by the predictions from the co-occurrence network and abundance patterns, acoel flatworms (*Symsagittifera* sp.) together with their photosynthetic green microalgal endosymbionts (*Tetraselmis* sp.) were collected in microplankton samples from *Tara* Oceans Station 22 in the Mediterranean Sea. Pictures show a 3D reconstructed specimen from LSCM images [green channel, cellular membranes (DiOC6); blue channel, DNA and the nuclei (Hoechst33342); red channel: chlorophyll autofluorescence]. (**A**) Co-occurrence plot of *Symsagittifera*- and *Tetraselmis*-related OTUs along *Tara* Oceans stations, showing the relatively high abundance of the holobiont at Station 22. (**B**) Dorsal view of the entire acoel flatworm specimen (~300 μm). The epidermis (green) is completely covered with cilia and displays some pore holes. (**C**) The removal of the green channel reveals the widespread distribution of small unicellular algae (red areas) inside the acoel body. The worm's nuclei display a clear signal (compact round blue shapes), whereas the algal nuclei are dimmer. A dinoflagellate theca (arrowhead) is located in the central syncytium, likely indicating predation. (**D**) Cross-section along a *z*-*y* plane allows localization of the algae, beneath the epidermis in the parenchyma. Only the external cell layer (green signal) from the dorsal view is visible because of the thickness and opacity of the worm. Scale bar, 50 μm.

are available at www.raeslab.org/companion/ocean-interactome.html.

Random forest-based models

Eukaryotic, prokaryotic, and environmental matrices were merged into two matrices [deep chlorophyll maximum layer (DCM) and surface water layer (SRF)]. For each of the three models [OTU versus other OTUs (M_{OTU}), environmental factors (M_{ENV}) or combined ($M_{OTU+ENV}$)], regressions were perfomed with OTU abundance as dependent and the abundances of other OTUs or environmental factors as independent variables. For each regression, up to 20 independent variables were selected by using the minimum Redundancy Maximum Relevance (mRMR) filter-ranking algorithm. Random forest regression (*25*) was followed by a leave-one-out cross-validation. The variable subset with the minimum leave-one-out NMSE (normalized mean square error) was selected. To identify the best model for a given target OTU, the significance of the NMSE difference was tested on the absolute error values [paired Wilcoxon test adjusted by Benjamini-Hochberg false discovery rate (FDR) estimation (*74*)]. NMSE computed on random data are larger than those from original data. In addition, M_{ENV} outperformed M_{OTU} when OTU abundances were randomized.

Variance partitioning

Environmental variables were *z* score–transformed; spatial variables (MEM eigenvectors) were calculated based on latitude and longitude (*75*). Forward selection (*76*) was carried out with function forward.sel in R-package packfor. Significance of the selected variables was assessed with 1000 permutations by using functions rda and anova.cca in vegan. Variance partitioning (*77*) was performed by using function varpart in vegan on Hellinger-transformed abundance data, the forward-selected environmental variables, and the forward-selected spatial variables and tested for significance with 1000 permutations.

Network inference

Taxon-taxon co-occurrence networks were constructed as in (*78*), selecting Spearman and Kullback-Leibler dissimilarity measures. To compute *P* values, we first generated permutation and bootstrap distributions, with 1000 iterations each, by shuffling taxon abundances and resampling from samples with replacement, respectively. The measure-specific *P* value was then obtained as the probability of the null value (represented by the mean of the permutation distribution) under a Gauss curve fitted to the mean and standard deviation of the bootstrap distribution. Permutations computed for Spearman included a renormalization step, which mitigates compositionality bias (ReBoot). Measure-specific *P* values were merged by using Brown's method (*79*) and multiple-testing-corrected with Benjamini-Hochberg (*74*). Last, edges with an adjusted *P* value above 0.05, with a score below the thresholds (*30*) or not supported by both measures after assessment of significance, were discarded.

Taxon-environment networks were computed with the same procedure, starting with 8000 initial positive and negative edges, each supported by both methods. For computational efficiency, we computed 23 taxon-taxon and taxon-environment networks separately, for two depths (DCM and SRF), four eukaryotic size fractions (0.8 to 5 μm, >0.8 μm, 20 to 180 μm, and 180 to 2000 μm) and their combinations, the prokaryotic size fraction (0.2 to 1.6 μm and 0.2 to 3.0 μm) and its combination with each of the eukaryotic and virus (<0.2 μm) size fractions. We then generated 23 taxon-environment union networks for environmental triplet detection and merged the taxon-taxon networks into a global network with 92,633 edges.

Estimation of false discovery rate

We estimated the FDR of network construction with two null models. The first shuffles counts while preserving overall taxon proportions and total sample count sums, but removing any dependencies between taxa. For the second,

we fitted a Dirichlet-multinomial distribution to the input matrix using the dirmult package in R (*80*) and generated a null matrix by sampling from this distribution, preserving total sample count sums. Null matrices were generated from count matrices (0.8 to 5 μm, 20 to 180 μm, and 180 to 2000 μm eukaryotic and prokaryotic size fraction as well as bacteriophage-prokaryotic composite, SRF, and DCM). Network construction was performed with the 20 null matrices and thresholds applied to the original matrices (*28*). From edge numbers in the original and the null networks, we estimated an average FDR of 9% (*28*).

Indirect taxon edge detection

For each taxon-environment union network, node triplets consisting of two taxa and one environmental parameter were identified. For each triplet, interaction information II was computed as $II = CI(X, Y \mid Z) - I(X, Y)$, where *CI* is the conditional mutual information between taxa *X* and *Y* given environmental parameter *Z*, and *I* is the mutual information between *X* and *Y*. *CI* and *I* were estimated by using minet (*81*). Taxon edges in environmental triplets were considered indirect when $II < 0$ and within the 0.05 quantile of the random II distribution obtained by shuffling environmental vectors (500 iterations). If a taxon pair was part of more than one environmental triplet, the triplet with minimum interaction information was selected.

For each environmental triplet, we also checked whether its sign pattern (the combination of positive and/or negative correlations) was consistent with an indirect interaction. From eight possible patterns, four indicate indirect relationships (for example, two negatively correlated taxa correlated with opposite signs to an environmental factor).

Network deconvolution (*32*) was carried out with $\beta = 0.9$. We considered an environmental triplet as indirect according to network deconvolution if any of its edges were removed.

All (11,043) negative interaction information triplets were consistent with an indirect relationship according to their sign patterns, and a majority (8209) was also supported by network deconvolution.

Influence of ocean regions on co-occurrence patterns

Samples were divided into groups according to region membership. The impact of each sample group on the Spearman correlation of each edge in the network was assessed by dividing the (absolute) omission score (OS) (Spearman correlation without these samples) by the absolute original Spearman score. To account for group size, the OS was computed repeatedly for random, same-sized sample sets. Nonparametric *P* values were calculated as the number of times random OSs were smaller than the sample group OS, divided by number of random OS (500 for each taxon pair). Edges were classified as region-specific when the ratio of OS and absolute original score was below 1 and multiple-testing-corrected *P* values (Benjamini-Hochberg) were below 0.05.

Overrepresentation analysis

Significance of taxon–taxon counts at high taxonomic ranks was assessed with the hypergeometric distribution implemented in the R function phyper. Mutual exclusion versus copresence analysis was performed by using the binomial distribution implemented in the R function pbinom, with the background probability estimated by the frequency of edges in the network.

Oceanic region analysis was also assessed by use of R's pbinom function, with the background probability estimated by dividing total ocean-specific edge number by total edge number. The *P* value was computed as the probability of obtaining the observed number of ocean-specific edges among the edges of a taxon pair. The same procedure was repeated for each oceanic region separately, with region-specific success probabilities. Edges classified as indirect were discarded before the analysis.

In all tests, *P* values were adjusted for multiple testing according to Benjamini, Hochberg, and Yekutieli (BY), implemented in the R function p.adjust.

Extracting functional groups from the global plankton interactome

Functional groups consist of a mix of major monophyletic lineages of parasites, together with classical polyphyletic PFTs, as defined in (*10*, *54*, *55*). Metabarcodes in the network were sorted into 15 parasite groups and seven PFTs (*55*) according to their (i) taxonomical classification, (ii) membership in a given size fraction, (iii) trophic mode, and (iv) biogeochemical role in dimethyl sulfide (DMS) production or silicification. After mapping the metabarcodes and their edges onto PFTs and parasites, edges are weighted by the number of links they represent. Overrepresentation of the number of links included in each edge was assessed with the hypergeometric distribution.

Parasite links in large fractions may point to parasite-host connections. We extracted all edges in the large fractions (20 to 180 μm and 180 to 2000 μm) between barcodes annotated as parasites and nonparasitic barcodes. Partners of parasites comprised potential hosts (Fig. 3B) but also organisms that are either too small or without size information. The former may represent unknown parasites (for example, coinfecting a host with known parasites), whereas the latter may represent previously unknown hosts.

Nestedness and modularity analysis

The analysis was carried out for 1869 positively correlated phage-prokaryotic pairs. Modularity was computed with the LP (Label propagation) BRIM algorithm (*82*) in BiMAT (*83*) with 100 permutations. Nestedness of the host-phage network as quantified with the NODF (nestedness with overlap and decreasing fill) algorithm (*84*) in BiMAT with 100 permutations (preserving edge number and degree distribution) was significant, but not with the NTC algorithm (*85*). We also tested the impact of random removal or addition of 5, 10, 15, and 20% edges. After random addition/deletion of edges, modularity and nestedness (according to NODF) remained significant.

Confirmation of predicted viruses-host associations

Two different approaches were used to confirm virus-host associations predicted by the co-occurrence network. First, the network host prediction was compared with the "known" host for viral populations closely related to an isolated virus—populations with more than 50% of predicted genes affiliated to the same phage reference genome [based on a BLASTp against RefseqVirus, threshold of 10^{-03} on e-value and 50 on bit score (*18*)]. Known phages corresponded to viruses infecting SAR11, SAR116, and Cyanobacteria, so that a predicted host was considered correct if affiliated to Alphaproteobacteria, Alphaproteobacteria, and Cyanobacteria, respectively [the lowest rank for which there was taxonomic assignment for those bacterial OTUs (*69*)]. This procedure was repeated on 1000 randomized networks (with same-degree distribution) to calculate the significance of the results. Second, contigs of putative hosts predicted by co-occurrence analysis were compared with BLAST to a set of viral sequences detected in draft and single-cell genomes with VirSorter (https://pods.iplantcollaborative.org/wiki/display/DEapps/VIRSorter+1.0.2). One contig (36DCM_3902) (Fig. 3E) displayed significant sequence similarity (blastn *e*-value < 10^{-151} over two segments) to one contig detected in a single-cell genome (AA160P02DRAFT_ scaffold_31.32). In order to compare the putative host associated to each contig, rRNA genes were predicted in the single-cell amplified genome (SAG) contigs with meta-rRNA (*86*). Sequences were annotated based on BLAST against the nonredundant (nr) database, and the comparison plot was generated with Easyfig (*87*).

Literature-based evaluation of predicted protist interactions

A panel of four experts, two specialized in the study of planktonic mutualistic protists (C.d.V. and J.D.) and two specialized in the study of planktonic parasitic protists (C. Berney and N.H.), screened literature looking for symbiotic interactions occurring among eukaryotic plankton. From this search, they built a list of 574 known symbiotic interactions sensu lato (parasitism and mutualism, at least one protist partner) in marine eukaryotic plankton, covering 197 eukaryotic genera, described in 76 publications since 1971. The experts extracted only symbiotic interaction cases described either from direct observation of both interacting partners through microscope (45%), sequence from symbiont isolated from the observed host (14%), or both (41%). Direct observation of partners interacting (86%) provides high confidence for the interaction, and the symbiont sequence allows its taxonomic identification. The protocol to build the list was the following: (i) the experts manually screened 3170 publications associated to each PR2 db sequence http://ssu-rrna.org/pr2 (*73*); (ii) the experts screened 293 publications

retrieved from Web of Science with the following query: "TOPIC:(plankton* AND (marin* OR ocean*)) AND (parasit* OR symbios* OR mutualis*)"; (iii) the experts screened GenBank 18*S* rDNA sequences of symbionts for which the "host" field was known. They labeled these interactions as "Unpublished." Last, the experts discussed any observed discordance until agreement was reached. The final table of literature-curated interactions includes a column indicating the type of evidence gathered about the interaction: 1 for only getting symbiont sequence, 2 for direct observation, and 3 for both. Symbiont GenBank host field belongs to category 1.

Experimental validation of a predicted interaction

V9 pairs were searched for organisms of suitable size in order to allow its isolation from morphological samples. This way, we targeted a predicted photosymbiosis between an acoel flatworm [V9 rDNA metabarcode 83% similar to *Symsagittifera psammophila* (*88*)] and a photosynthetic microalga (*Tara* Oceans V9 metabarcode 100% similar to a *Tetraselmis* sp) (*89*).

Fifteen acoel specimens (hosts) were isolated from formaldehyde-4% microplankton samples of station 22 (A100000458), in which both partner OTUs displayed high abundances. Before imaging, specimens were rinsed with artificial seawater, then DNA and membrane structures were stained for 60 min with 10 μM Hoechst 33342 and 1.4 μM DiOC6(3) (Life Technologies, Grand Island, NY). Microscopy was conducted by using a Leica TCS SP8 (Leica Microsystems, Wetzlar, Germany) confocal laser scanning microscope and a HC PL APO 40x/1.10 W motCORR CS2 objective. The DiOC6 signal (ex488nm/em500-520nm) was collected simultaneously with the chlorophyll signal (ex488nm/em670-710nm), followed by the Hoechst signal (ex405 nm/em420-470nm). Images were processed with Fiji (*90*), and 3D specimens were reconstructed with Imaris (Bitplane, Belfast, UK).

To obtain the sequences of the metabarcodes of each partner, seven acoels were isolated from ethanol-preserved samples from station 22 (TARA_A100000451), individually rinsed in filtered seawater, and stored at −20°C in absolute ethanol. DNA was extracted with MasterPureTM DNA/RNA purification kit (Epicenter, Madison, WI) and polymerase chain reaction amplified by using the universal-eukaryote primers (forward 1389F and reverse 1510R) from (*10*). Chlorophyte-specific primers (Chloro2F: 5′- CGTATATTTAAGTT-GYTGCAG-3′ and Tetra2-rev 5′- CAGCAATGGGC-GGTGGC GAAC-3′) were designed to amplify the microalgae V9 rDNA as in (*4*). Purified amplicons were subjected to poly-A reaction and ligated in pCR®4-TOPO TA Cloning vector (Invitrogen, Carlsbad, CA), cloned by using chemically competent *Escherichia coli* cells, and Sanger-sequenced with the ABI-PRISM Big Dye Terminator Sequencing kit (Applied Biosystems, Foster City, CA) by using the 3130xl Genetic Analyzer (Applied Biosystems).

REFERENCES AND NOTES

1. F. Azam *et al*., The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.* **10**, 257–263 (1983). doi: 10.3354/meps010257
2. A. W. Thompson *et al*., Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science* **337**, 1546–1550 (2012). doi: 10.1126/science.1222700; pmid: 22997339
3. A. Chambouvet, P. Morin, D. Marie, L. Guillou, Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science* **322**, 1254–1257 (2008). doi: 10.1126/science.1164387; pmid: 19023082
4. J. Decelle *et al*., An original mode of symbiosis in open ocean plankton. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18000–18005 (2012). doi: 10.1073/pnas.1212303109; pmid: 23071304
5. V. Smetacek, Making sense of ocean biota: How evolution and biodiversity of land organisms differ from that of the plankton. *J. Biosci.* **37**, 589–607 (2012). doi: 10.1007/s12038-012-9240-4; pmid: 22922185
6. J. L. Sabo, L. R. Gerber, "Trophic ecology," *AccessScience* (McGraw-Hill Education, 2014); available at www.accessscience.com/content/trophic-ecology/711650.
7. F. Rohwer, R. V. Thurber, Viruses manipulate the marine environment. *Nature* **459**, 207–212 (2009). doi: 10.1038/nature08060; pmid: 19444207
8. P. G. Verity, V. Smetacek, Organism life cycles, predation, and the structure of marine pelagic ecosystems. *Mar. Ecol. Prog. Ser.* **130**, 277–293 (1996). doi: 10.3354/meps130277
9. A. Z. Worden *et al*., Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science* **347**, 1257594 (2015). doi: 10.1126/science.1257594; pmid: 25678667
10. C. de Vargas *et al*., *Science* **348**, XXX–XXX (2014).
11. K. Faust, J. Raes, Microbial interactions: From networks to models. *Nat. Rev. Microbiol.* **10**, 538–550 (2012). doi: 10.1038/nrmicro2832; pmid: 22796884
12. S. Chaffron, H. Rehrauer, J. Pernthaler, C. von Mering, A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res.* **20**, 947–959 (2010). doi: 10.1101/gr.104521.109; pmid: 20458099
13. J. Raes, I. Letunic, T. Yamada, L. J. Jensen, P. Bork, Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Mol. Syst. Biol.* **7**, 473 (2011). doi: 10.1038/msb.2011.6; pmid: 21407210
14. J. A. Gilbert *et al*., Defining seasonal marine microbial community dynamics. *ISME J.* **6**, 298–308 (2012). doi: 10.1038/ismej.2011.107; pmid: 21850055
15. J. M. Beman, J. A. Steele, J. A. Fuhrman, Co-occurrence patterns for abundant marine archaeal and bacterial lineages in the deep chlorophyll maximum of coastal California. *ISME J.* **5**, 1077–1085 (2011). doi: 10.1038/ismej.2010.204; pmid: 21228895
16. C.-E. T. Chow, D. Y. Kim, R. Sachdeva, D. A. Caron, J. A. Fuhrman, Top-down controls on bacterial community structure: Microbial network analysis of bacteria, T4-like viruses and protists. *ISME J.* **8**, 816–829 (2014). doi: 10.1038/ismej.2013.199; pmid: 24196323
17. E. Karsenti *et al*., A holistic approach to marine eco-systems biology. *PLOS Biol.* **9**, e1001177 (2011). doi: 10.1371/journal.pbio.1001177; pmid: 22028628
18. J. R. Brum *et al*., *Science* **348**, XXX–XXX (2014).
19. S. Sunagawa *et al*., *Science* **348**, XXX–XXX (2014).
20. E. Villar *et al*., *Science* **348**, XXX–XXX (2014).
21. Companion web site table w2; available at www.raeslab.org/companion/ocean_interactome/tables/W2.xlsx.
22. A. Meot, P. Legendre, D. Borcard, *Environ. Ecol. Stat.* **5**, 1–27 (1998). doi: 10.1023/A:1009693501830
23. Companion web site table w3; available at www.raeslab.org/companion/ocean_interactome/tables/W3.xlsx.
24. Companion web site figure w1; available at www.raeslab.org/companion/ocean_interactome/figures//W1.pdf .
25. L. Breiman, *Mach. Learn.* **45**, 5–32 (2001). doi: 10.1023/A:1010933404324
26. Companion web site table w4; available at www.raeslab.org/companion/ocean_interactome/tables/W4.xlsx.
27. Companion web site figure w2; available at http://www.raeslab.org/companion/ocean_interactome/figures/W2.pdf.
28. Companion web site table w5; available at www.raeslab.org/companion/ocean_interactome/tables/W5.xlsx.
29. Companion web site table w6; available at http://www.raeslab.org/companion/ocean_interactome/tables/W6.xlsx
30. Companion web site table w7; available at www.raeslab.org/companion/ocean_interactome/tables/W7.xlsx.
31. Companion web site figure w3; available at http://www.raeslab.org/companion/ocean_interactome/figures/W3.pdf.
32. S. Feizi, D. Marbach, M. Médard, M. Kellis, Network deconvolution as a general method to distinguish direct dependencies in networks. *Nat. Biotechnol.* **31**, 726–733 (2013). doi: 10.1038/nbt.2635; pmid: 23851448
33. Companion web site table w8 available at www.raeslab.org/companion/ocean_interactome/tables/W8.xlsx.
34. M. E. Cusick *et al*., Literature-curated protein interaction datasets. *Nat. Methods* **6**, 39–46 (2009). doi: 10.1038/nmeth.1284; pmid: 19116613
35. Companion web site table w9; available at www.raeslab.org/companion/ocean_interactome/tables/W9.xlsx.
36. J. A. Fuhrman, J. A. Cram, D. M. Needham, Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* **13**, 133–146 (2015). doi: 10.1038/nrmicro3417; pmid: 25659323
37. Companion web site additional material is available at www.raeslab.org/companion/ocean_interactome/Accompanying_Material.docx.
38. Companion web site figure w4; available at www.raeslab.org/companion/ocean_interactome/figures/W4.pdf.
39. Companion web site table w10; available at www.raeslab.org/companion/ocean_interactome/tables/W10.xlsx.
40. G. B. McManus, L. F. Santoferrara, *The Biology and Ecology of Tintinnid Ciliates* (John Wiley & Sons, New York, 2012.
41. J. Cachon, in *Ann sci nat b*. (Paris, 1964), vol. 6, p. 1.
42. P. S. Salomon, E. Granéli, M. H. C. B. Neves, E. G. Rodriguez, Infection by Amoebophrya spp. parasitoids of dinoflagellates in a tropical marine coastal area. *Aquat. Microb. Ecol.* **55**, 143–153 (2009). doi: 10.3354/ame01293
43. F. Gómez, P. López-García, A. Nowaczyk, D. Moreira, The crustacean parasites Ellobiopsis Caullery, 1910 and Thalassomyces Niezabitowski, 1913 form a monophyletic divergent clade within the Alveolata. *Syst. Parasitol.* **74**, 65–74 (2009). doi: 10.1007/s11230-009-9199-1; pmid: 19633933
44. S. Ohtsuka *et al*., Morphology and host-specificity of the apostome ciliate Vampyrophrya pelagica infecting pelagic copepods in the Seto Inland Sea, Japan. *Mar. Ecol. Prog. Ser.* **282**, 129–142 (2004). doi: 10.3354/meps282129
45. A. Skovgaard, S. A. Karpov, L. Guillou, The parasitic dinoflagellates Blastodinium spp. inhabiting the gut of marine, planktonic copepods: Morphology, ecology, and unrecognized species diversity. *Front. Microbiol.* **3**, 305 (2012). doi: 10.3389/fmicb.2012.00305; pmid: 22973263
46. L. Stemmann *et al*., Global zoogeography of fragile macrozooplankton in the upper 100-1000 m inferred from the underwater video profiler. *ICES J. Mar. Sci.* **65**, 433–442 (2008). doi: 10.1093/icesjms/fsn010
47. J. M. Gasol, P. A. Del Giorgio, C. M. Duarte, *Biomass Distribution in Marine Planktonic Communities* (American Society of Limnology and Oceanography, Waco, TX, 1997), vol. 42.
48. B. A. Ward, S. Dutkiewicz, M. J. Follows, Modelling spatial and temporal patterns in size-structured marine plankton communities: Top-down and bottom-up controls. *J. Plankton Res.* **36**, 31–47 (2014). doi: 10.1093/plankt/fbt097
49. A. Ianora *et al*., Aldehyde suppression of copepod recruitment in blooms of a ubiquitous planktonic diatom. *Nature* **429**, 403–407 (2004). doi: 10.1038/nature02526; pmid: 15164060
50. M. Martinez-Garcia *et al*., Unveiling in situ interactions between marine protists and bacteria through single cell sequencing. *ISME J.* **6**, 703–707 (2012). doi: 10.1038/ismej.2011.126; pmid: 21938022
51. E. T. Jolley, A. K. Jones, The interaction between *Navicula muralis* grunow and an associated species of *Flavobacterium*. *Br. Phycol. J* **12**, 315–328 (1977). doi: 10.1080/00071617700650341
52. T. R. Miller, R. Belas, Dimethylsulfoniopropionate metabolism by Pfiesteria-associated Roseobacter spp. *Appl. Environ. Microbiol.* **70**, 3383–3391 (2004). doi: 10.1128/AEM.70.6.3383-3391.2004; pmid: 15184135
53. Companion web site figure w5; available at www.raeslab.org/companion/ocean_interactome/figures/W5.pdf.
54. C. Le Quere *et al*., *Glob. Change Biol.* **11**, 17 (2005).
55. Companion web site table w11; available at www.raeslab.org/companion/ocean_interactome/tables/W11.xlsx.
56. A. Skovgaard, *Acta Protozool.* **53**, 51 (2014).
57. Companion web site figure w6; available at www.raeslab.org/companion/ocean_interactome/figures/W6.pdf.
58. S. von der Heyden, E. E. Chao, K. Vickerman, T. Cavalier-Smith, Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of euglenozoa. *J. Eukaryot. Microbiol.* **51**, 402–416 (2004). doi: 10.1111/j.1550-7408.2004.tb00387.x; pmid: 15352322

59. F. Gómez, D. Moreira, K. Benzerara, P. López-García, Solenicola setigera is the first characterized member of the abundant and cosmopolitan uncultured marine stramenopile group MAST-3. *Environ. Microbiol.* **13**, 193–202 (2011). doi: 10.1111/j.1462-2920.2010.02320.x; pmid: 20722698
60. C. E. Hamm *et al*., Architecture and material properties of diatom shells provide effective mechanical protection. *Nature* **421**, 841–843 (2003). doi: 10.1038/nature01416; pmid: 12594512
61. P. Assmy, V. Smetacek, in *Encyclopedia of Microbiology*, M. Schaechter, Ed. (Elsevier, Oxford, UK, 2009), pp. 27–41.
62. Companion web site table w12; available at www.raeslab.org/companion/ocean_interactome/tables/W12.xlsx.
63. Companion web site figure w7; available at www.raeslab.org/companion/ocean_interactome/figures/W7.pdf.
64. J. S. Weitz *et al*., Phage-bacteria infection networks. *Trends Microbiol.* **21**, 82–91 (2013). doi: 10.1016/j.tim.2012.11.003; pmid: 23245704
65. Companion web site table w13; available at www.raeslab.org/companion/ocean_interactome/tablesW13.xlsx.
66. Companion web site figure w9; available at www.raeslab.org/companion/ocean_interactome/companion_figures/W9.pdf.
67. S. Pesant *et al*., Open science resources for the discovery and analysis of *Tara* Oceans data. http://biorxiv.org/content/early/2015/05/08/019117 (2015).
68. R. Logares *et al*., Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ. Microbiol.* **16**, 2659–2671 (2013).
69. E. Pruesse *et al*., SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007). doi: 10.1093/nar/gkm864; pmid: 17947321
70. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010). doi: 10.1093/bioinformatics/btq461; pmid: 20709691
71. Q. Wang, G. M. Garrity, J. M. Tiedje, J. R. Cole, Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007). doi: 10.1128/AEM.00062-07; pmid: 17586664
72. F. Mahé, T. Rognes, C. Quince, C. de Vargas, M. Dunthorn, Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ* **2**, e593 (2014). doi: 10.7717/peerj.593; pmid: 25276506
73. L. Guillou *et al*., The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41** (D1), D597–D604 (2013). doi: 10.1093/nar/gks1160; pmid: 23193267
74. Y. Benjamini, Y. Hochberg, *J. R. Stat. Soc., B* **57**, 289 (1995).
75. D. Borcard, P. Legendre, C. Avois-Jacquet, H. Tuomisto, DIssecting the spatial structure of ecological data at multiple scales. *Ecology* **85**, 1826–1832 (2004). doi: 10.1890/03-3111
76. F. G. Blanchet, P. Legendre, D. Borcard, Forward selection of explanatory variables. *Ecology* **89**, 2623–2632 (2008). doi: 10.1890/07-0986.1; pmid: 18831183
77. D. Borcard, P. Legendre, P. Drapeau, Partialling out the spatial component of ecological variation. *Ecology* **73**, 1045 (1992). doi: 10.2307/1940179
78. K. Faust *et al*., Microbial co-occurrence relationships in the human microbiome. *PLOS Comput. Biol.* **8**, e1002606 (2012). doi: 10.1371/journal.pcbi.1002606; pmid: 22807668
79. M. B. Brown, 400: A method for combining non-independent, one-sided tests of significance. *Biometrics* **31**, 987 (1975). doi: 10.2307/2529826
80. T. Tvedebrink, Overdispersion in allelic counts and θ-correction in forensic genetics. *Theor. Popul. Biol.* **78**, 200–210 (2010). doi: 10.1016/j.tpb.2010.07.002; pmid: 20633572
81. P. E. Meyer, F. Lafitte, G. Bontempi, minet: A R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics* **9**, 461 (2008). doi: 10.1186/1471-2105-9-461; pmid: 18959772
82. L. Xin, T. Murata, in *Web Intelligence and Intelligent Agent Technologies, 2009* (WI-IAT 09. IEEE/WIC/ACM International Joint Conferences, 2009), vol. 1, pp. 50.
83. C. O. Flores, T. Poisot, S. Valverde, J. S. Weitz, http://arxiv.org/abs/1406.6732 (2014)
84. M. J. Barber, Modularity and community detection in bipartite networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **76**, 066102 (2007). doi: 10.1103/PhysRevE.76.066102; pmid: 18233893
85. W. Atmar, B. D. Patterson, The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* **96**, 373–382 (1993). doi: 10.1007/BF00317508
86. Y. Huang, P. Gilna, W. Li, Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* **25**, 1338–1340 (2009). doi: 10.1093/bioinformatics/btp161; pmid: 19346323
87. M. J. Sullivan, N. K. Petty, S. A. Beatson, Easyfig: A genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011). doi: 10.1093/bioinformatics/btr039; pmid: 21278367
88. I. Ruiz-Trillo, M. Riutort, D. T. J. Littlewood, E. A. Herniou, J. Baguña, Acoel flatworms: Earliest extant bilaterian Metazoans, not members of Platyhelminthes. *Science* **283**, 1919–1923 (1999). doi: 10.1126/science.283.5409.1919; pmid: 10082465
89. R. J. Gast, T. A. McDonnell, D. A. Caron, srDna-based taxonomic affinities of algal symbionts from a planktonic foraminifer and a solitary radiolarian. *J. Phycol.* **36**, 172–177 (2000). doi: 10.1046/j.1529-8817.2000.99133.x
90. J. Schindelin *et al*., Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012). doi: 10.1038/nmeth.2019; pmid: 22743772
91. M. E. Newman, Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **74**, 036104 (2006). doi: 10.1103/PhysRevE.74.036104; pmid: 17025705

ACKNOWLEDGMENTS

We thank the commitment of the following people and sponsors: Centre National de la Recherche Scientifique (CNRS) (in particular, Groupement de Recherche GDR3280), European Molecular Biology Laboratory (EMBL), Genoscope/CEA, VIB, Stazione Zoologica Anton Dohrn, UNIMIB, Fund for Scientific Research – Flanders (G.L.M., K.F., S.C., and J.R.), Rega Institute (J.R.), KU Leuven (J.R.), The French Ministry of Research, the French Government "Investissements d'Avenir" programmes OCEANOMICS (ANR-11-BTBR-0008), FRANCE GENOMIQUE (ANR-10-INBS-09-08), MEMO LIFE (ANR-10-LABX-54), PSL* Research University (ANR-11-IDEX-0001-02), ANR (projects POSEIDON/ANR-09-BLAN-0348, PHYTBACK/ANR-2010-1709-01, PROMETHEUS/ANR-09-PCS-GENM-217, TARA GIRUS/ANR-09-PCS-GENM-218, European Union FP7 (MicroB3/No.287589, IHMS/HEALTH-F4-2010-261376, ERC Advanced Grant Awards to CB (Diatomite: 294823), Gordon and Betty Moore Foundation grant (3790) to M.B.S., Spanish Ministry of Science and Innovation grant CGL2011-26848/BOS MicroOcean PANGENOMICS to S.G.A., TANIT (CONES 2010-0036) from the Agència de Gestió d´Ajusts Universitaris i Reserca funded to S.G.A., JSPS KAKENHI grant number 26430184 to H.O., FWO, BIO5, Biosphere 2, Agnès b., the Veolia Environment Foundation, Region Bretagne, Lorient Agglomeration, World Courier, Illumina, the EDF Foundation, FRB, the Prince Albert II de Monaco Foundation, Etienne Bourgois, the *Tara* schooner, and its captain and crew. We are also grateful to the French Ministry of Foreign Affairs for supporting the expedition and to the countries that graciously granted sampling permissions. *Tara* Oceans would not exist without continuous support from 23 institutes (http://oceans.taraexpeditions.org). We also acknowledge the EMBL Advanced Light Microscopy Facility (ALMF), and in particular R. Pepperkok. The authors further declare that all data reported herein are fully and freely available from the date of publication, with no restrictions, and that all of the samples, analyses, publications, and ownership of data are free from legal entanglement or restriction of any sort by the various nations whose waters the *Tara* Oceans expedition sampled in. Data described herein is available at www.raeslab.org/companion/ocean-interactome.html, at the EBI under the projects PRJEB402 and PRJEB6610, and at Pangaea http://doi.pangaea.de/10.1594/PANGAEA.840721, http://doi.pangaea.de/10.1594/PANGAEA.840718, http://doi.pangaea.de/10.1594/PANGAEA.843018 and http://doi.pangaea.de/10.1594/PANGAEA.843022 and on table S1. The data release policy regarding future public release of *Tara* Oceans data are described in Pesant *et al*. (*67*). All authors approved the final manuscript. This article is contribution number 25 of Tara Oceans. The supplementary materials contain additional data.

TARA OCEANS COORDINATORS

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

[1]Department of Marine Biology and Oceanography, Institute of Marine Sciences ICM-CSIC, Barcelona, Spain. [2]Structural and Computational Biology, European Molecular Biology Laboratory, Heidelberg, Germany. [3]School of Marine Sciences, University of Maine, Orono, USA. [4]Environmental and Evolutionary Genomics Section, Institut de Biologie de l'Ecole Normale Supérieure, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8197, Institut National de la Santé et de la Recherche Médicale U1024, Ecole Normale Supérieure, Paris, France. [5]CNRS, UMR 7144, Station Biologique de Roscoff, Roscoff, France. [6]Sorbonne Universités, UPMC Univ Paris 06, UMR 7144, Station Biologique de Roscoff, Roscoff, France. [7]Dept of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, USA. [8]CNRS, UMR 7093, Laboratoire d'Océanographie de Villefranche (LOV), Observatoire Océanologique, F-06230 Villefranche-sur-mer, France. [9]Sorbonne Universités, UPMC Paris 06, UMR 7093, Laboratoire d'Océanographie de Villefranche (LOV), Observatoire Océanologique, F-06230 Villefranche-sur-mer, France. [10]CNRS UMR 7232, BIOM, Banyuls-sur-Mer, France. [11]Sorbonne Universités, OOB, UPMC Paris 06, Banyuls-sur-Mer, France. [12]Aix Marseille Université, CNRS, IGS UMR 7256, Marseille, France. [13]Laboratory of Ecology and Evolution of Plankton, Stazione Zoologica Anton Dohrn, Naples, Italy. [14]CEA, Genoscope, Evry France. [15]CNRS, UMR 8030, Evry, France. [16]Université d'Evry, UMR 8030, Evry, France. [17]Environmental and Evolutionary Genomics Section, Institut de Biologie de l'Ecole Normale Supérieure, CNRS, UMR 8197, Institut National de la Santé et de la Recherche Médicale U1024, Ecole Normale Supérieure, Paris, France. [18]Directors' Research, European Molecular Biology Laboratory, Heidelberg, Germany. [19]Cell Biology and Biophysics, European Molecular Biology Laboratory, Heidelberg, Germany. [20]Institute for Chemical Research, Kyoto University, Kyoto, Japan. [21]PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, Bremen, Germany. [22]MARUM, Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany. [23]Department of Microbiology and Immunology, Rega Institute KU Leuven, Leuven, Belgium. [24]VIB Center for the Biology of Disease, VIB, Leuven, Belgium. [25]Laboratory of Microbiology, Vrije Universiteit Brussel, Brussels, Belgium. [26]School of Biology and Environmental Science, University College Dublin, Dublin, Ireland. [27]Bigelow Laboratory for Ocean Sciences, East Boothbay, USA. [28]Department of Geosciences, Laboratoire de Météorologie Dynamique (LMD), Ecole Normale Supérieure, Paris, France. [29]Laboratoire de Physique des Océan, UBO-IUEM, Polouzané, France. [30]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, USA. [31]DVIP Consulting, Sèvres, France.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6237/1262073/suppl/DC1
Table S1

3 October 2014; accepted 18 March 2015
10.1126/science.1262073

Determinants of community structure in the global plankton interactome

Gipsi Lima-Mendez *et al.*

Science **348**, (2015);
DOI: 10.1126/science.1262073