



# The planktonic protist interactome: where do we stand after a century of research?

Marit F. Markussen Bjorbækmo<sup>1</sup> · Andreas Evenstad<sup>1</sup> · Line Lieblein Røsæg<sup>1</sup> · Anders K. Krabberød<sup>1</sup> · Ramiro Logares<sup>1,2</sup>

Received: 14 May 2019 / Revised: 17 September 2019 / Accepted: 24 September 2019 / Published online: 4 November 2019  
© The Author(s) 2019. This article is published with open access

## Abstract

Microbial interactions are crucial for Earth ecosystem function, but our knowledge about them is limited and has so far mainly existed as scattered records. Here, we have surveyed the literature involving planktonic protist interactions and gathered the information in a manually curated *Protist Interaction Database* (PIDA). In total, we have registered ~2500 ecological interactions from ~500 publications, spanning the last 150 years. All major protistan lineages were involved in interactions as hosts, symbionts (mutualists and commensalists), parasites, predators, and/or prey. Predation was the most common interaction (39% of all records), followed by symbiosis (29%), parasitism (18%), and ‘unresolved interactions’ (14%, where it is uncertain whether the interaction is beneficial or antagonistic). Using bipartite networks, we found that protist predators seem to be ‘multivorous’ while parasite–host and symbiont–host interactions appear to have moderate degrees of specialization. The SAR supergroup (i.e., Stramenopiles, Alveolata, and Rhizaria) heavily dominated PIDA, and comparisons against a global-ocean molecular survey (*TARA Oceans*) indicated that several SAR lineages, which are abundant and diverse in the marine realm, were underrepresented among the recorded interactions. Despite historical biases, our work not only unveils large-scale eco-evolutionary trends in the protist interactome, but it also constitutes an expandable resource to investigate protist interactions and to test hypotheses deriving from omics tools.

## Introduction

Aquatic microbes, including unicellular eukaryotes and prokaryotes, are essential for the functioning of the biosphere [1–4]. Microbes exist in diverse ecological

communities where they interact with each other as well as with larger multicellular organisms and viruses.

Interaction between microbial species has played important roles in evolution and speciation. One of the best examples is that the origin of eukaryotes is grounded in the interaction-events of endosymbiosis; giving rise to mitochondria, chloroplasts, and other metabolic capacities in the eukaryotic cell [5–8]. Microbial interactions guarantee ecosystem function, having crucial roles in, for instance, carbon channeling in photosymbiosis, control of microalgae blooms by parasites, and phytoplankton-associated bacteria influencing the growth and health of their host. Despite their importance, our understanding of microbial interactions in the ocean and other aquatic systems is rudimentary, and the majority of them are still unknown [4, 9–11]. The earliest surveys of interactions between aquatic microbes date back to the 19th century. In 1851, while on board H.M.S *Rattlesnake* in the Pacific Ocean, Thomas Huxley discovered small yellow–green cells inside the conspicuous planktonic radiolarians which he thought were organelles [12]. Later on, Karl Brandt established that the yellowish cells were symbiotic alga

---

These authors contributed equally: Marit F. Markussen Bjorbækmo, Andreas Evenstad, Line Lieblein Røsæg

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41396-019-0542-5>) contains supplementary material, which is available to authorized users.

- ✉ Anders K. Krabberød  
a.k.krabberod@ibv.uio.no
- ✉ Ramiro Logares  
ramiro.logares@icm.csic.es

<sup>1</sup> Department of Biosciences, Section for Genetics and Evolutionary Biology (Evogene), University of Oslo, Blindernv. 31, N-0316 Oslo, Norway

<sup>2</sup> Institut de Ciències del Mar (CSIC), Passeig Marítim de la Barceloneta, 37-49, ES-08003 Barcelona, Catalonia, Spain

and named them *Zooxanthella nutricula* [13]. Since these early studies, hundreds of others have reported microbial interactions by using classic tools, mainly microscopy, but this knowledge has not yet been gathered into one accessible database. Over the last ~15 years, High-Throughput Sequencing (HTS) [14–16] of environmental DNA or RNA has transformed our understanding of microbial diversity [17] and evolution [18]. Furthermore, HTS studies have generated hypotheses on microbial interactions based on correlations of estimated microbial abundances over spatiotemporal scales [19–22]. These hypotheses need to be tested with other types of data, such as known interactions from the literature [23]. Overall, HTS will allow to start addressing key driving questions in microbiology such as, what are the main types of interactions in the protist world? Does cooperation outweigh competition among protists? What is the architecture of the protist interactome? And how does this interactome change over spatiotemporal scales?

Here, our main objectives were to assemble the knowledge on aquatic protist interactions from the literature and make it available to the scientific community. We also report the main patterns found in this survey. We examined the available scientific literature spanning the last ~150 years, and recorded ~2500 ecological interactions from ~500 publications going back to the late 1800s [24] (Supplementary Fig. 1). Based on this, we generated a manually curated and publicly available *Protist Interaction DAtabase* (PIDA; <https://doi.org/10.5281/zenodo.1195514>). PIDA entries have been grouped into four types of pairwise ecological interactions: *parasitism*, *predation*, *symbiosis*, and ‘*unresolved interaction*’. Parasitism is an antagonistic relationship between organisms, which is beneficial to one partner but harmful to the other, while predation refers for the most part to the engulfment of smaller cells through phagocytosis. In PIDA symbiosis refers to interactions beneficial for both partners (mutualism, e.g., photosymbiosis) or beneficial for one and potentially neutral for the other (commensalism, e.g., host defense). The fourth category ‘unresolved interactions’ are associations where it is uncertain whether the interactions are beneficial or antagonistic to the involved partners. The taxonomic classification in PIDA includes genus and species level, in addition to three levels that were chosen pragmatically to make the database user-friendly and portable.

## Materials and methods

PIDA was assembled between January and November 2017 through a recursive survey of papers on microbial interactions published between 1894 and 2017. The

search strategy to find the relevant literature and the template for organizing the database was performed following Lima-Mendez et al. [21]. Initially, reviews resulting from the Boolean search string (plankton\* AND (marin\* OR ocean\*)) AND (parasit\* OR symbios\* OR mutualis\*) in Scopus (<https://www.scopus.com/>) and Web of Science (<http://webofknowledge.com/>) were examined, then the references therein were further explored. In addition, literature on protist predation on other protists and bacteria were also screened. Entries from the *AquaSymbio* database (<http://aquasymbio.fr/>) were compared against the entries in PIDA, and occasionally used as a source of additional literature. The overlap between *AquaSymbio* and PIDA is ~20% (~500 entries). Many of the entries in *AquaSymbio* are interactions between protists and multicellular organisms, therefore they are not included in PIDA.

PIDA documents the ecological interaction between two organisms, identified down to the species level, if possible. Interactions are characterized as *parasitism*, *predation*, *symbiosis* (either mutualism or commensalism), or ‘*unresolved*’. Parasitism is used in cases where the study clearly identifies a parasitic interaction. Cases of kleptoplasty and mixotrophy together with classical predation are contained within the group of entries termed predation. Symbiosis includes endo- and ectosymbiosis and is categorized into the different forms of symbiosis (e.g., photosymbiosis). The unresolved interactions include associations/interactions between organisms where it is yet unknown whether the associations are beneficial or antagonistic.

In addition to genus and species levels, the taxonomic classification includes three additional levels chosen pragmatically to make the database more user-friendly and portable. The highest level distinguishes between eukaryotes and prokaryotes. The second level places each taxon within supergroups or other high taxonomic ranks (e.g., Rhizaria or Alveolata) following the scheme of Adl et al. [25, 26]. The third level places each taxon in groups below the supergroup taxonomic rank (phylum, e.g., Ciliophora, Dinoflagellata, and Acantharia, or class levels, e.g., Chlorophyceae, Kinetoplastea, and Diplomonadida). The taxonomic names at the third level follows the nomenclature of the SILVA database (release 128, May/June 2017) [27–29]. Species names in PIDA have been updated to the most recent agreed-upon classification and can therefore deviate from the original papers they stem from due to synonymization. PIDA also documents the methods used to determine the interacting species. Symbionts and/or hosts determined by any form of microscopy or direct observation are denoted (1). Symbionts and/or hosts determined by sequencing or Fluorescence In Situ Hybridization (FISH) are denoted (2). The combination of the former two is denoted (3). Most interactions with observation type 2 also

have GenBank [30] accession numbers. A published paper is associated to each interaction entry, and when a DOI is available, it is included. Only interactions from aquatic systems are included (marine, brackish, and freshwater). The resulting PIDA contains 2422 entries from 528 publications and is publicly available at github (<https://github.com/ramalok/PIDA>).

## Bipartite networks

Bipartite networks are the representation of interactions between two distinct classes of nodes, such as plant–pollinator, parasite–host, or prey–predator. Identifying patterns in bipartite networks is useful in explaining their formation and function. We investigated how symbiosis, parasitism, and predation differ in terms of specialization. For example, if parasite taxa have a broader host range compared with the host range of symbionts, this indicates that parasites are less specialized (and consequently more generalists) than symbionts. We also used the bipartite networks to investigate whether predators are omnivorous (generalists) or picky (specialists) in their diets. All analyses were conducted in the statistical environment R v. 3.5.0 [31]. We constructed bipartite qualitative (binary) directional networks using the R-package *bipartite* v. 2.08 [32]. All taxa where the taxonomy assigned to one of the ‘partners’ in PIDA was ‘unknown eukaryote’, ‘unidentified bacteria’, or ‘unidentified prokaryote’ were removed before further analyses of the bipartite networks. Bipartite network indices were calculated using the functions *networklevel* [33] and *specieslevel* [34] (default settings except `weighted = FALSE`) in the R-package *bipartite*. Bipartite networks and network analyses were performed at four taxonomic levels (‘supergroup’, ‘phylum’, genus, and species). Found patterns were consistent across the taxonomic levels, therefore only the species level is shown. Degree (number of links/edges/interactions per node) was calculated for prey, predators, parasites, symbionts, ‘interactors’, and hosts at the species level. The specialization index  $d'$  (Kullback–Leibler distance) [35], measures the degree of specialization at the species level, and was calculated as deviation of the actual interaction frequencies from a null model that assumes all partners in the other level of the bipartite network are used in proportion to their availability. The specialization index  $d'$  ranges from 0 for the most generalist to 1 for the most specialist, and was calculated for prey, predators, parasites, symbionts, ‘interactors’, and hosts at the species level.

All barplots and density plots were constructed using the R-package *ggplot2* v. 3.1.0 [36] and the networks in Fig. 1a–d and Supplementary Fig. 2 were visualized in Cytoscape v. 3.6.1 [32, 37].

## Interlinked species

Interlinked species are taxa that are present in either several types of interactions, or on both sides of the same interaction. Interlinked species were determined using the R-package *systemPipeR* v. 3.8 [38] for all interaction types, except the unresolved interactions (since the nature of these interaction is unknown). Only taxa with full species names were included to avoid overestimating overlapping species (e.g., *Amoebophrya* sp. and similar were excluded from the list of overlapping species). Venn diagram intersects were computed using the function *overLapper* and plotted using the function *vennPlot*. Parasites only overlapped with parasite hosts and were subsequently added to the Venn plot.

## Results

### Aquatic microbial interactions

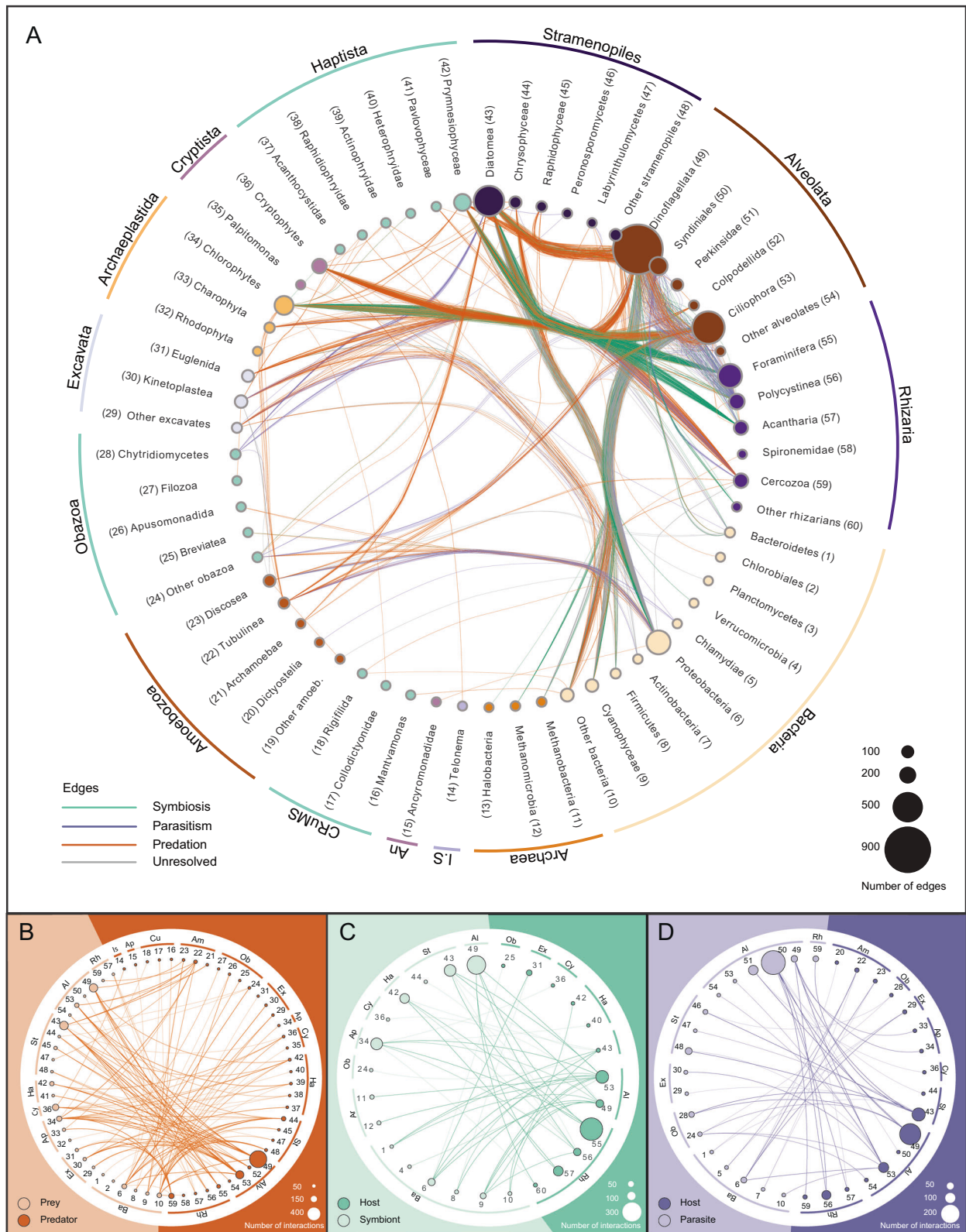
The literature in PIDA was dominated by studies based on direct observation of interactions such as light microscopy. In total, 82% of the entries were based on microscopy, and only 38% of those were combined with molecular methods. The most commonly studied interaction in the literature was predation, representing 39% of all entries, followed by symbiosis (29%), parasitism (18%), and unresolved interactions (14%).

The SAR supergroup (Alveolata, Stramenopiles, and Rhizaria) dominated with ~92% of the total entries (Figs. 1 and 2). Of all host and predator records, ~90% belonged to the SAR supergroup (Alveolata 51%, Stramenopiles 12%, and Rhizaria 27%; Fig. 2). The SAR supergroup was less dominant as symbiont/ parasite/ interactor/ prey, but still represented the largest group, with 50% of all entries (Alveolata 33%, Stramenopiles 16%, and Rhizaria 1%; Fig. 2).

The majority of interactions (82%) were from marine or brackish waters, while studies from freshwater systems accounted for a smaller fraction of the interactions (18%). This is not surprising given the larger number of studies from the marine phototrophic zone compared to other environments.

### Predator–prey interactions

Predator–prey interactions constituted the majority of entries in PIDA. We acknowledge that separating predation into categories represents a simplified version of how predator–prey interactions are in nature, where e.g., mixotrophy is challenging old paradigms [39–42]. However, for simplicity, we here divided predation into three different



categories depending on the type of prey involved: herbivory (grazing on autotrophic/chloroplast containing eukaryotic algae), bacterivory (feeding on autotrophic and/

or heterotrophic bacteria) or ‘carnivory’ (predation on other heterotrophic protists). Based on this definition, herbivory was the most common form of predation in our survey (68%



◀ **Fig. 1** Overview of the interactions included in PIDA. Nomenclature and taxonomic order of Eukaryota is based on Adl et al. 2019 [25]. Nomenclature and taxonomic order of Bacteria is based on Schultz et al. 2017 [87]. The nodes are grouped (outer circle) according to eukaryotic supergroups (or *Incertae sedis*), Bacteria and Archaea. **a** Network based on the 2422 entries in PIDA. Nodes represent eukaryotic and prokaryotic taxa and are colored accordingly. Node size indicates the number of edges/links that are connected to that node. Each node/taxon is assigned a number, which corresponds with the numbers for taxa in **b–d**. Edges represent interactions between two taxa and are colored according to ecological interaction type: *predation* (orange), *symbiosis* (green), and *parasitism* (purple). The network is undirected, meaning that a node can contain both parasites/symbionts/prey and hosts/predators. To avoid cluttering of the figure, ‘Self-loops’, which represent cases where both interacting organisms belong to the same taxon (e.g., a dinoflagellate eating another dinoflagellate) are not shown as edges/links in this figure, but are considered in the size of nodes. The outermost circle groups taxa in the different eukaryotic ‘supergroups’ or the prokaryotic domains Bacteria and Archaea. *Ancryomonadidae* is abbreviated An. *Telonema* is not placed into any of the supergroups, but classified as *Incertae sedis* (abbreviated *I.S.* in the figure). In **b**, **c**, and **d** the following abbreviations for supergroups are used: Ar Archaea, Ba Bacteria, Rh Rhizaria, Al Alveolata, St Stramenopiles, Ha Haptista, Cy Cryptista, Ap Archaeplastida, Ex Excavata, Ob Obazoa, Am Amoebozoa, Cu CRuMS, An *Ancryomonadidae*, Is *Incertae sedis*. **b** Predator–prey interactions in PIDA. The node numbers correspond to taxa node numbers in **a**. Abbreviations for supergroups are described above. Background and nodes are colored according to functional role in the interaction: Prey are colored light orange (left part of figure), while predators are depicted in dark orange (right part of figure). The size of each node represents the number of edges connected to that node. **c** Symbiont–host interactions included in PIDA. The node numbers correspond to node numbers in **a**. Abbreviations for supergroups are described above. Symbionts are to the left, colored light green, and their hosts are to the right in dark green. The size of each node represents the number of edges connected to that node. **d** Parasite–host interactions included in PIDA. The node numbers correspond to node numbers in **a**. Abbreviations for supergroups are described above. Parasite taxa are depicted in light purple (left), hosts in dark purple (right). (The unresolved interactions are shown in Supplementary Fig. 7)

of the predator–prey interactions) with all the major eukaryotic lineages represented among the predators. Entries of herbivore dinoflagellates and ciliates (both Alveolata) dominated (Fig. 1a, b). Bacterivory accounted for 16% of the predator–prey interactions and was also documented in most eukaryotic groups (Fig. 1b), with an expected predominance of small heterotrophic flagellates.

### Symbiont–host interactions

Symbiotic *protist–protist* interactions made up 81% of the symbiont entries and all of these interactions represented photosymbiosis. Dinoflagellates, diatoms, chlorophytes, trebouxiphyceans, and prymnesiophytes accounted for most of the recorded photosymbionts, living in symbiosis with rhizarian, ciliate, and dinoflagellate hosts (Fig. 1a, c). *Bacteria–protist* interactions represented 16% of the total

number of symbiont entries in PIDA, and was dominated by bacterial entries belonging to Proteobacteria and Cyanophyceae that mainly interacted with Alveolata (dinoflagellates and ciliates), Stramenopiles (diatoms), and Excavata (euglenids); Fig. 1a, c. The bacteria–protist interactions were involved in many different types of symbiotic relationships, from photosymbiosis (13%) to nitrogen fixation (46%) and vitamin exchange (36%). Symbiotic *archaea–protist* interactions represented 3% of symbiont entries in PIDA, and the majority of these were methanogenic symbiont interactions between archaeal Metanomicrobia and anaerobic Ciliophora (Fig. 1a, c).

### Unresolved interactions

The unresolved interactions represent all ecological interactions where the functional role of the relationship between the partners was not determined. Several of these cases likely represent commensalism. The unresolved ‘interactor’–host category mainly consisted of *protist–bacteria* interactions (73%), dominated by interactions between Proteobacteria and Alveolata (ciliates and dinoflagellates) or Proteobacteria and Stramenopiles (diatoms; Fig. 1a & Supplementary Fig. 2). *Protist–protist* interactions represented 27% of the unresolved interactions, and mainly included alveolate, excavate, and stramenopile symbionts that interacted with alveolate, rhizarian, stramenopile, and amoebozoan hosts (Fig. 1a & Supplementary Fig. 2). The amoebozoan hosts (*Neoparamoeba* spp.) were only registered to interact with unknown kinetoplastids (Excavata), which is likely an example of an unusual form of endosymbiosis [43]. The biological nature of these interactions still remains unknown.

### Parasite–host interactions

Parasites in PIDA were dominated by a few taxonomic groups that all belonged to Alveolata, such as Syndiniales (~50% *Amoebophrya*), Perkinsidae (~98% *Parvilucifera*), and Dinoflagellata. Together they accounted for 2/3 of the parasite entries (Fig. 1a, d). These alveolate parasites mainly infected other alveolates such as dinoflagellates and ciliates, but rhizarian and diatom hosts were also recorded (Fig. 1d). Parasites belonging to different stramenopiles lineages such as Peronosporomycetes (oomycetes), Labyrinthulomycetes, and *Pirsonia* were mainly described from diatom hosts (Fig. 1d). Rhizarian parasites constituted 5% of the parasite records and were represented by just a few cercozoans and phagomyxids, which parasitized diatoms, as well as the rhizarian phytomyxid *Woronina pythii*, which parasitized different *Pythium* species (Peronosporomycetes). Parasitic fungi from Chytridiomycetes, Microsporidia, and Sordariomycetes (the last two included in ‘other Obazoa’ in

**Fig. 2** Interactions in PIDA. Number of interactions registered in the PIDA database for the different taxonomic groups at ‘supergroup level’ (corresponding to the second taxonomic level in PIDA). Red bars show predation, purple parasitism, green represent symbiosis, and gray unresolved interactions. Solid colors represent predator/host and transparent colors represent prey/symbiont/parasite/interactor. Because CRuMS, Ancyromonadidae, and *Incertae sedis* comprised very few entries (5, 1, and 2 predator entries, respectively), they are not included in this figure

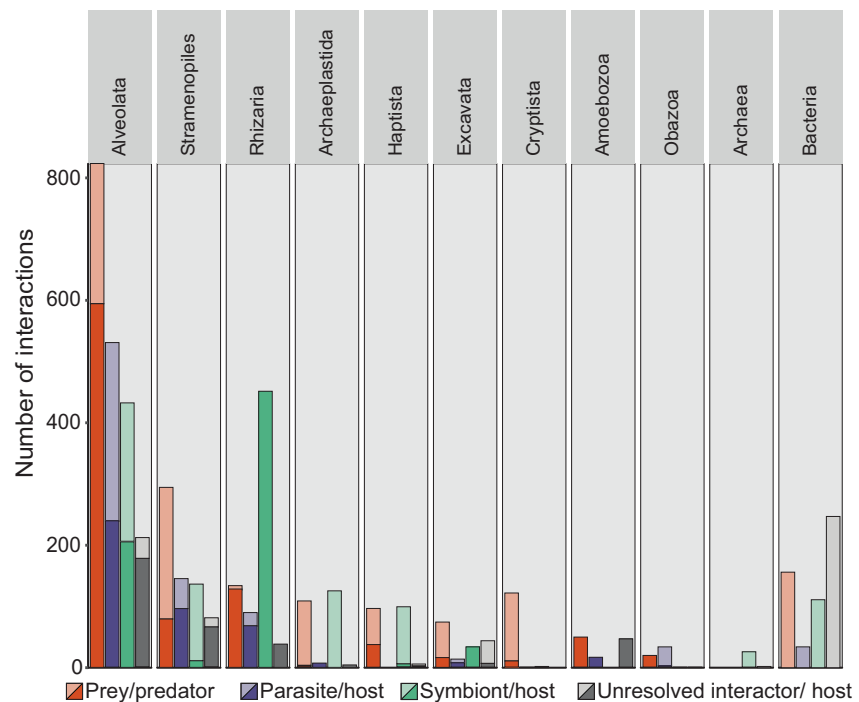


Fig. 1a, d) were also represented by relatively few entries (only 7% of the parasite records). Yet, the records of parasitic fungi demonstrated that they infect a relatively broad range of protists, such as dinoflagellates, apicomplexans, ciliates, and diatoms. Bacterial parasites of protists accounted for 8% of the parasite entries and were registered mainly from amoebozoan, excavate, and ciliate hosts (Fig. 1d).

### Bipartite interaction networks

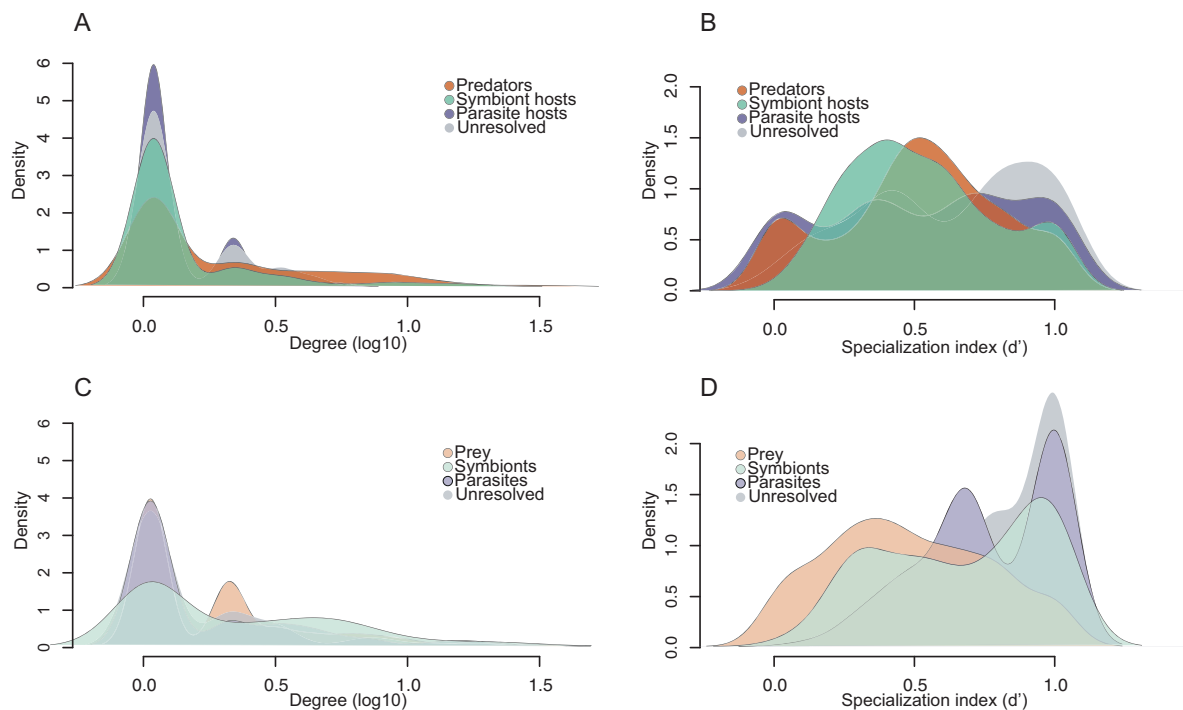
Since PIDA consists of pairwise interactions between aquatic microbes where the roles of the participants are known we can represent the interactions as bipartite networks. Bipartite networks provide a systematic way of representing data that consist of two distinct guilds, such as plant–pollinator, parasite–host, or predator–prey. These networks are composed of nodes (representing species or genera) connected by links (edges) representing the interactions between nodes. The *degree* of a node (species) is the sum of links connecting the particular node to the nodes from the other guild. Consequently, a higher degree value indicates a higher level of generalism [33]. For example, a parasite that has gone through multiple host-shifts and has the capacity to parasitize different hosts would display a higher degree than a parasite specialized to interact with only one host. We have constructed binary (presence/absence) bipartite networks for predator–prey, symbiont–host, and parasite–host interactions, as well as for the unresolved interactions (Supplementary Figs. 3–6). We calculated specialization indices to analyze variation in

specialization within the bipartite networks and to examine if the four interaction types differed in terms of specialization (Fig. 3a–d; Supplementary Fig. 7; Table 1).

The *predator–prey* bipartite interaction networks had 342 prey and 337 predator species (Supplementary Fig. 3). Although the number of prey and predators in the network were almost equal, there were multiple shared interactions. That is, several predators feed on the same prey (i.e., the prey has a high degree) and conversely, generalist predators preying on multiple prey organisms (i.e., predators having a high degree; Fig. 3a, c; Supplementary Figs. 3 and 7). The five predators with highest degree included four dinoflagellates and one cercozoan (Table 1).

The prey organisms with highest degree belonged to Haptista, Cryptista, Stramenopiles, and Alveolata (Table 1). The specialization index ( $d'$ ) was uniformly distributed from 0 (generalist) to 1 (specialist) indicating that predation was not driving predator or prey to specialization (Fig. 3b, d).

The *symbiont–host* interaction networks consisted of 98 symbiont and 188 host species (Supplementary Fig. 4). The majority of both hosts and symbionts had a low degree (Fig. 3a, c; Supplementary Fig. 7). The distribution of the specialization index ( $d'$ ) for hosts indicates that PIDA includes both, specialists interacting with only one or a few symbionts, as well as hosts that interact with multiple symbionts (Fig. 3b). The five host taxa that had the highest number of associated symbionts (i.e., highest degree) were four foraminiferans (Rhizaria) and one ciliate (Alveolata; Table 1). Very few hosts were ‘true generalists’ (i.e., with  $d'$  close to 0, Fig. 3b). The symbionts in PIDA had high  $d'$



**Fig. 3** Density plots for degree and specialization indices for the bipartite networks in PIDA. **a** Degree (number of links/edges/interactions) for predators, hosts of parasites, hosts of symbionts, and hosts of interactors in the bipartite networks. **b** Specialization index  $d'$  (Kullback–Leibler distance) [35], for predators, hosts of parasites,

hosts of symbionts, and host of interactors. The specialization index  $d'$  ranges from 0 for the most generalized to 1 for the most specialized. **c** Degree (number of links/edges/interactions) for prey, parasites, symbionts, and interactors. **d** Specialization index  $d'$  for prey, parasites, symbionts, and interactors

values in general, which indicate that they are specialized ( $d'$  0.75–1; Fig. 3d). This was coherent with the low degree of most symbionts, which showed few links to different hosts (Fig. 3c; Supplementary Fig. 7). The five symbionts with highest degree included taxa belonging to five different ‘supergroups’ (Haptista, Stramenopiles, Cyanobacteria, Alveolata, and Chlorophyta; Table 1).

The network for the *unresolved interactor–host* interactions had 85 ‘interactors’ and 141 host species (Supplementary Fig. 5). The distribution of the specialization index ( $d'$ ) for hosts and interactors showed high  $d'$  values for most taxa (Fig. 3b, d). Concordantly, the majority of hosts and interactors also had a low degree (Fig. 3a, c; Supplementary Fig. 7). Altogether, this could indicate that the majority of the *unresolved interactor–host* interactions are specialized, or that they are simply understudied. The five interactors and the five hosts with highest degree are shown in Table 1.

The network for *parasitism* had 130 parasites and 262 hosts (Supplementary Fig. 6). Hosts were dominated by taxa with low degree (i.e., few parasites per host, maximum number was five; Table 1), which indicated that they are infected by a relatively low number of parasites (Fig. 3a; Supplementary Fig. 7). The  $d'$  values showed, however, that there was an equal distribution of host taxa ranging from ‘true generalists’ ( $d'$  value of 0) to ‘true specialists’ ( $d'$  value of 1; Fig. 3b). The parasites had for the most part a

low degree, and the distribution of the specialization index indicated that several of the parasites were specialists ( $d'$  values ~1; Fig. 3d). Parasites showed the highest relative number of specialized taxa in PIDA. However, parasites also included the taxa with the highest degree (Fig. 3c; Supplementary Fig. 7), the well-known parasites belonging to Syndiniales (MALV II) and Perkinsidae (Table 1).

### Interlinked species

Interlinked species [44] are taxa present in either several types of interactions, or on both sides of the same interaction. An interlinked species is for example a species that is registered as a predator in the predator–prey network, and is also present as a host in the symbiont–host network. The unresolved interactions were excluded from these analyses since the nature of these interactions are unknown. In total there were 94 interlinked species in PIDA (~4% of the total entries, Fig. 4; Table 2). The maximum number of interaction types for any species was three (Table 2; panels A–D). The majority of interlinked species occurred in the overlap of species recorded as predator, as prey and as host of parasites (Table 2; panel A). There was only a single interlinked species that held a role in each of the three independent bipartite networks (i.e., predator–prey, symbiont–host, and parasite–host), *Paramecium bursaria* (Table 2, panel B; corresponding to the

**Table 1** Degree and specialization index ( $d'$ ), for the five species with highest degree (highest number of links/edges) in the bipartite networks for each interaction type

Type	Taxa <sup>a</sup>	Species	Phylum (Supergroup)	Degree	$d'$
Predators	337	<i>Leptophrys vorax</i>	Cercozoa (Rhizaria)	34	0.8
		<i>Oblea rotunda</i>	Dinoflagellata (Alveolata)	27	0.6
		<i>Karodinium armiger</i>	Dinoflagellata (Alveolata)	26	0.6
		<i>Diplopsalis lenticula</i>	Dinoflagellata (Alveolata)	15	0.4
		<i>Diplopsalopsis bomba</i>	Dinoflagellata (Alveolata)	15	0.4
Hosts of symbionts	188	<i>Amphistegina lobifera</i>	Foraminifera (Rhizaria)	17	0.6
		<i>Amphistegina lessonii</i>	Foraminifera (Rhizaria)	15	0.6
		<i>Borelis schlumbergi</i>	Foraminifera (Rhizaria)	12	0.6
		<i>Heterostegina depressa</i>	Foraminifera (Rhizaria)	12	0.5
		<i>Paramecium bursaria</i>	Ciliophora (Alveolata)	9	0.8
Hosts of interactors	141	<i>Diophrys scutum</i>	Ciliophora (Alveolata)	4	0.2
		<i>Euplotes harpa</i>	Ciliophora (Alveolata)	4	0.9
		<i>Euplotes woodruffi</i>	Ciliophora (Alveolata)	4	0.7
		<i>Hemigastrostyla elongata</i>	Ciliophora (Alveolata)	4	0.3
		<i>Petalomonas spagnophila</i>	Euglenida (Excavata)	4	0.4
Hosts of parasites	262	<i>Alexandrium minutum</i>	Dinoflagellata (Alveolata)	5	0.3
		<i>Collozoum</i> NA	Polycystinea (Rhizaria)	5	0.8
		<i>Guinardia delicatula</i>	Diatomea (Stramenopiles)	5	0.6
		<i>Alexandrium catenella</i>	Dinoflagellata (Alveolata)	4	0.1
		<i>Coscinodiscus granii</i>	Diatomea (Stramenopiles)	4	0.5
Prey	342	<i>Isochrysis galbana</i>	Prymnesiophyceae (Haptista)	31	0.3
		<i>Rhodomonas salina</i>	Cryptophyceae (Cryptista)	27	0.4
		<i>Skeletonema costatum</i>	Diatomea (Stramenopiles)	27	0.3
		<i>Heterocapsa triquetra</i>	Dinoflagellata (Alveolata)	20	0.3
		<i>Heterosigma akashiwo</i>	Raphidiophyceae (Stramenopiles)	19	0.2
Symbionts	98	<i>Phaeocystis</i> NA	Prymnesiophyceae (Haptista)	29	1.0
		diatom NA	Diatomea (Stramenopiles)	21	0.9
		cyanophyte NA	Cyanophyceae (Cyanobacteria)	18	0.9
		dinoflagellate NA	Dinoflagellata (Alveolata)	15	0.9
		Chlorella NA	Trebouxiophyceae (Chlorophyta)	11	0.7
Interactors	85	alphaproteobacter NA	Alphaproteobacteria (Proteobacteria)	20	0.5
		gammaproteobacteria NA	Gammaproteobacteria (Proteobacteria)	19	0.6
		<i>Gyrodinium</i> NA	Dinoflagellata (Alveolata)	12	1.0
		cyanophyte NA	Cyanophyceae (Cyanobacteria)	8	1.0
		<i>Polynucleobacter necessarius</i>	Betaproteobacteria (Proteobacteria)	7	0.8
Parasites	130	<i>Amoebophrya</i> NA	Syndiniales (Alveolata)	61	0.8
		<i>Parvilucifera infectans</i>	Perkinsidae (Alveolata)	29	0.5
		<i>Euduboscquella</i> NA	Syndiniales (Alveolata)	18	1.0
		<i>Cryothecomonas longipes</i>	Cercozoa (Rhizaria)	14	0.6
		<i>Pirsonia formosa</i>	<i>I.s.</i> Stramenopiles	14	0.6

<sup>a</sup>Taxa displays the number of species registered in PIDA for the different interaction types. The taxonomy of the five species with highest degree is shown at species, phylum, and supergroup level. Degree shows the number of edges for the top five taxa. Specialization index  $d'$  (Kullback–Leibler distance) [30] ranges from 0 for the most generalized to 1 for the most specialized. Abbreviations used: *Type* interaction type.





**Table 2** Interlinked species in PIDA

Panel	Species	Phylum	Supergroup	No	Present as		
A	<i>Akashiwo sanguinea</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Alexandrium catenella</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Alexandrium ostenfeldii</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Alexandrium tamarense</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Cochlodinium polykrikoides</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Dinophysis acuminata</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Gymnodinium catenatum</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Gymnodinium sanguineum</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Heterocapsa rotundata</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Heterocapsa triquetra</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Neoceratium furca</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Oblea rotunda</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Oxyrrhis marina</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Prorocentrum micans</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Prorocentrum minimum</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Protoperidinium pellucidum</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
A	<i>Scrippsiella trochoidea</i>	Dinoflagellata	Alveolata	3	Prey	Predator	Host of par
B	<i>Paramecium bursaria</i>	Ciliophora	Alveolata	3	Prey	Host of symb	Host of par
C	<i>Heterocapsa rotunda</i>	Dinoflagellata	Alveolata	3	Symbiont	Prey	Predator
D	<i>Noctiluca scintillans</i>	Dinoflagellata	Alveolata	3	Predator	Host of symb	Host of par
D	<i>Thalassicola nucleata</i>	Polycystinea	Rhizaria	3	Predator	Host of symb	Host of par
E	<i>Durinskia baltica</i>	Dinoflagellata	Alveolata	2	Host of symb	Host of par	–
E	<i>Durinskia dybowskii</i>	Dinoflagellata	Alveolata	2	Host of symb	Host of par	–
E	<i>Kryptoperidinium foliaceum</i>	Dinoflagellata	Alveolata	2	Host of symb	Host of par	–
E	<i>Stentor polymorphus</i>	Ciliophora	Alveolata	2	Host of symb	Host of par	–
E	<i>Acanthometra pellucida</i>	Acantharia	Rhizaria	2	Host of symb	Host of par	–
E	<i>Acanthostaurus purpurascens</i>	Acantharia	Rhizaria	2	Host of symb	Host of par	–
E	<i>Collozoum caudatum</i>	Polycystinea	Rhizaria	2	Host of symb	Host of par	–
F	<i>Chlamydomonas hedleyi</i>	Chlorophyceae	Archaeplastida	2	Prey	Symbiont	–
F	<i>Nitzschia frustulum</i>	Diatomea	Stramenopiles	2	Prey	Symbiont	–
F	<i>Protodinium simplex</i>	Dinoflagellata	Alveolata	2	Prey	Symbiont	–
G	<i>Keppenodinium mycetoides</i>	Syndiniales	Alveolata	2	Parasite	Host of par	–
G	<i>Oodinium acanthometrae</i>	Dinoflagellata	Alveolata	2	Parasite	Host of par	–
H	<i>Acanthamoeba castellanii</i>	Discosea	Amoebozoa	2	Predator	Host of par	–
H	<i>Acanthamoeba polyphaga</i>	Discosea	Amoebozoa	2	Predator	Host of par	–
H	<i>Dinophysis acuta</i>	Dinoflagellata	Alveolata	2	Predator	Host of par	–
H	<i>Dinophysis norvegica</i>	Dinoflagellata	Alveolata	2	Predator	Host of par	–
H	<i>Diplopsalis lenticula</i>	Dinoflagellata	Alveolata	2	Predator	Host of par	–
H	<i>Euplotes woodruffi</i>	Ciliophora	Alveolata	2	Predator	Host of par	–
H	<i>Eutintinnus pectinis</i>	Ciliophora	Alveolata	2	Predator	Host of par	–
H	<i>Favella ehrenbergii</i>	Ciliophora	Alveolata	2	Predator	Host of par	–
H	<i>Gonyaulax polygramma</i>	Dinoflagellata	Alveolata	2	Predator	Host of par	–
H	<i>Karlodinium veneticum</i>	Dinoflagellata	Alveolata	2	Predator	Host of par	–
H	<i>Trithigmostoma cucullulus</i>	Ciliophora	Alveolata	2	Predator	Host of par	–
H	<i>Protoperidinium bipes</i>	Dinoflagellata	Alveolata	2	Predator	Host of par	–
H	<i>Protoperidinium minutum</i>	Dinoflagellata	Alveolata	2	Predator	Host of par	–
H	<i>Protoperidinium steinii</i>	Dinoflagellata	Alveolata	2	Predator	Host of par	–
H	<i>Strombidium capitatum</i>	Ciliophora	Alveolata	2	Predator	Host of par	–
I	<i>Coscinodiscus granii</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Coscinodiscus radiatus</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Cylindrotheca closterium</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Eucampia zoodiacus</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Chaetoceros didymus</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Guinardia delicatula</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Guinardia flaccida</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Leptocylindrus danicus</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–

**Table 2** (continued)

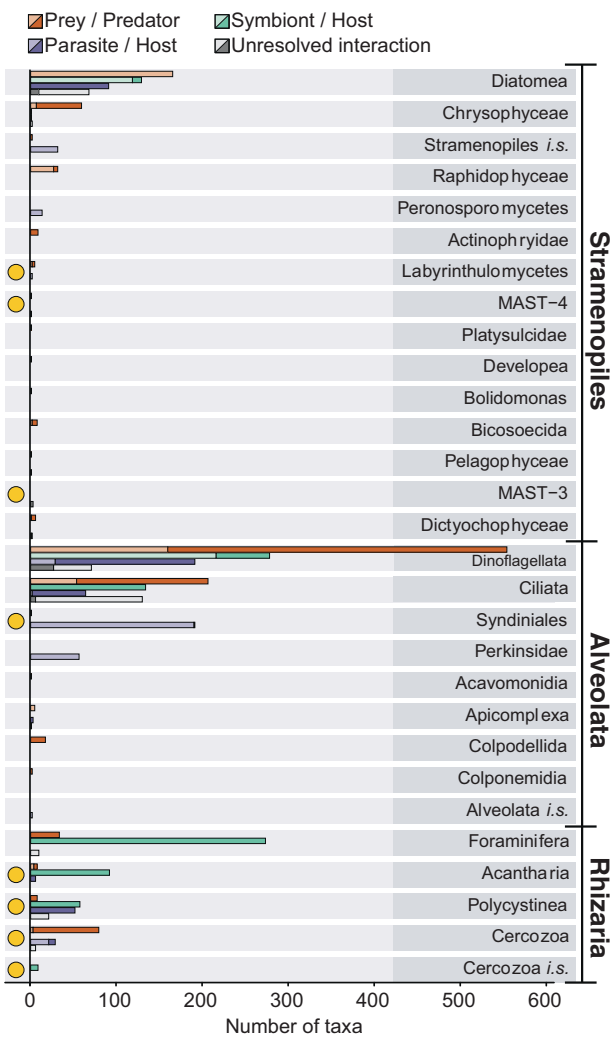
Panel	Species	Phylum	Supergroup	No	Present as		
I	<i>Stephanopyxis turris</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Thalassionema nitzschioides</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Thalassiosira nordenskiöldii</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Thalassiosira punctigera</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Thalassiosira rotula</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Odontella sinensis</i>	Diatomea	Stramenopiles	2	Prey	Host of par	–
I	<i>Gymnodinium instriatum</i>	Dinoflagellata	Alveolata	2	Prey	Host of par	–
I	<i>Gymnodinium mikimotoi</i>	Dinoflagellata	Alveolata	2	Prey	Host of par	–
I	<i>Gyrodinium aureolum</i>	Dinoflagellata	Alveolata	2	Prey	Host of par	–
I	<i>Neoceratium fusus</i>	Dinoflagellata	Alveolata	2	Prey	Host of par	–
I	<i>Neoceratium lineatum</i>	Dinoflagellata	Alveolata	2	Prey	Host of par	–
I	<i>Neoceratium tripos</i>	Dinoflagellata	Alveolata	2	Prey	Host of par	–
I	<i>Paramecium tetraurelia</i>	Ciliophora	Alveolata	2	Prey	Host of par	–
J	<i>Thalassiosira pseudonana</i>	Diatomea	Stramenopiles	2	Prey	Host of symb	–
K	<i>Heterosigma akashiwo</i>	Raphidophyceae	Stramenopiles	2	Prey	Predator	–
K	<i>Pseudobodo tremulans</i>	Bicosoecida	Stramenopiles	2	Prey	Predator	–
K	<i>Fibrocapsa japonica</i>	Raphidophyceae	Stramenopiles	2	Prey	Predator	–
K	<i>Archerella flavum</i>	Labyrinthulomycetes	Stramenopiles	2	Prey	Predator	–
K	<i>Lingulodinium polyedrum</i>	Dinoflagellata	Alveolata	2	Prey	Predator	–
K	<i>Myrionecta rubra</i>	Ciliophora	Alveolata	2	Prey	Predator	–
K	<i>Pfiesteria piscicida</i>	Dinoflagellata	Alveolata	2	Prey	Predator	–
K	<i>Prorocentrum triestinum</i>	Dinoflagellata	Alveolata	2	Prey	Predator	–
K	<i>Euplotes aediculatus</i>	Ciliophora	Alveolata	2	Prey	Predator	–
K	<i>Favella taraikaensis</i>	Ciliophora	Alveolata	2	Prey	Predator	–
K	<i>Amphidinium carterae</i>	Dinoflagellata	Alveolata	2	Prey	Predator	–
K	<i>Rhynchomonas nasuta</i>	Kinetoplastea	Excavata	2	Prey	Predator	–
K	<i>Bodo saltans</i>	Kinetoplastea	Excavata	2	Prey	Predator	–
K	<i>Prymnesium parvum</i>	Prymnesiophyceae	Haptophyta	2	Prey	Predator	–
K	<i>Chrysochromulina polylepis</i>	Prymnesiophyceae	Haptophyta	2	Prey	Predator	–
L	<i>Peridinium quinquecorne</i>	Dinoflagellata	Alveolata	2	Predator	Host of symb	–
L	<i>Euplotes daidaleos</i>	Ciliophora	Alveolata	2	Predator	Host of symb	–
L	<i>Climacostomum virens</i>	Ciliophora	Alveolata	2	Predator	Host of symb	–
L	<i>Orbulina universa</i>	Foraminifera	Rhizaria	2	Predator	Host of symb	–
L	<i>Archaias angulatus</i>	Foraminifera	Rhizaria	2	Predator	Host of symb	–
L	<i>Globigerinoides ruber</i>	Foraminifera	Rhizaria	2	Predator	Host of symb	–
L	<i>Globigerinoides sacculifer</i>	Foraminifera	Rhizaria	2	Predator	Host of symb	–
L	<i>Globorotalia menardii</i>	Foraminifera	Rhizaria	2	Predator	Host of symb	–
L	<i>Lenisia limosa</i>	Breviatea	Obazoa	2	Predator	Host of symb	–

Panels A–L refer to the 12 overlapping sections in the Venn diagram (Fig. 5). The taxonomy of the overlapping species in PIDA is listed at species, phylum and supergroup level. The column ‘No’ refers to the number of roles the overlapping species held, i.e., the number of bipartite networks (or interaction types in the same network) the species occurred in. The column ‘Present as’ displays which interaction types, or networks the overlapping species occurred in (i.e., what roles they held). Only taxa with full species name determined were included to avoid overestimating overlapping species (e.g., *Amoebophrya* sp. and similar were excluded). Abbreviations: host of par = host of parasite, host of symb = Host of symbiont.

improved with future studies and observations. In particular, future interconnections between PIDA and molecular studies should enhance the power of both and lead to more solid conclusions. Despite its biases, PIDA can be used to initialize biological hypotheses that should be further tested, and can serve as a guide for future experiments on interactions in underrepresented protists groups.

Our comprehensive bibliographic survey shows that microbial interactions are spread across all major eukaryotic groups as well as across the main bacterial (e.g., Cyanobacteria, Bacteroidetes, Proteobacteria, and Firmicutes) and

archaeal (Halobacteria, Methanobacteria, and Methanomicrobia) lineages. All major protistan groups were involved in interactions in one or multiple ways; as hosts, symbionts, parasites, predators, and/or prey (Fig. 1a–d). The well-known representatives of the SAR supergroup (i.e., Stramenopiles, Alveolata, and Rhizaria) were however overrepresented compared with the other ‘supergroups’ in PIDA. Within SAR the distribution was further skewed toward certain well-characterized and species-rich lineages, such as dinoflagellates, ciliates, and diatoms. SAR members have historically gained much attention since they hold



**Fig. 5** The dominating SAR supergroup. Number of interactions registered in PIDA belonging to the SAR supergroup (Stramenopiles, Alveolates, and Rhizaria). For each of the SAR supergroups the entries of the different taxonomic groups at ‘phylum level’ are shown (corresponding to the third taxonomic level in PIDA). Red bars show predators, green represent symbiosis, purple parasitism, and gray unresolved interactions. Solid colors represent host/predator and transparent colors represent prey/symbiont/parasite/interactor. Yellow circles highlight the ‘phyla’ that comprise few records in PIDA compared with the (hyper)diversity these ‘phyla’ represent in environmental HTS studies, such as the *Tara Oceans* study [45]. Abbreviations used in this figure: *i.s.* refers to *Incertae sedis* or unknown

important roles as primary producers, parasites, symbionts, and predators/grazers [49]). In particular, the dinoflagellate and diatom lineages include many species that hold key roles as primary producers, fixing substantial amounts of atmospheric CO<sub>2</sub> in the surface ocean [50], and have consequently been the subject of many studies. Dinoflagellates also include species that can form harmful algal blooms (HABs) [51], producing toxins that can have devastating effects on fisheries and aquaculture, making the research on their ecology and life cycles a priority [52]. Other

heterotrophic dinoflagellates and ciliates are among the most important microzooplankton predators in the ocean, consuming between 60 and 80% of the primary production every day at a global scale [53–56]. Although the SAR supergroup dominated PIDA, our comparison of the SAR records with the *Tara Oceans* data [45] revealed that there are several SAR lineages, which are abundant and diverse in the marine realm, that were underrepresented when it comes to characterization of their ecological roles as interactors in aquatic environments, such as Labyrinthulomycetes and MAST (Stramenopiles), Syndiniales (MALV II; Alveolata), Acantharia, Polycystinea, and Cercozoa (Rhizaria).

Compared with environmental studies there were also several other protist lineages that were underrepresented in PIDA. Fungi, for example, have been shown to be diverse in environmental HTS surveys of several aquatic environments [57–59] and some of these have been proposed as important parasites of protists [60, 61]. Judging from the scant number of entries in PIDA and the relatively broad host ranges that these organisms had, we suggest that future investigations should focus on revealing more about the ecological function of aquatic fungi. Another underrepresented lineage in PIDA is Excavata. In the *Tara Oceans* surveys the diplomonids were shown to be hyperdiverse, with more than 12 000 OTUs, while they have only four entries in PIDA. It was not surprising that diplomonids were poorly represented in PIDA since their immense diversity was only recently discovered [45], and because little is known about the lifestyle of these excavates [62]. But it underlines that diplomonids and other excavates represent a black box also when it comes to ecological interactions.

Protist predation or grazing is crucial for channeling carbon and energy to higher trophic levels [55, 63] as well as for the release of dissolved nutrients to the base of aquatic food webs [64]. The bipartite network analysis of the predator–prey interactions in PIDA indicated that predation as an ecological strategy is not directed toward either specialization nor generalization, but that predators are ‘multivorous’ and feed on several different prey organisms. Instead of hunting for prey many organisms depend on other strategies for resource acquisition that involve more intimate relationships, such as the interaction between parasites or symbionts and their hosts. These intimate interactions have evolved independently from freeliving ancestors multiple times in diverse and evolutionary unrelated protist lineages. To develop such intimate interactions requires a high degree of specialization as it necessitates a metabolic dialog between the interacting organisms. Our bipartite network analyses for symbiont–host and parasite–host interactions showed that many of the symbiont and parasite species seemed to be moderately specialized.



The importance and scientific relevance of symbiosis was reflected by the great variety of symbiotic interaction we found. Several protists harbor microbial symbionts (eukaryotic and prokaryotic) that provide e.g., carbohydrates (through photosynthesis), vitamins, Nitrogen (through  $N_2$ -fixation), and defense to their hosts, in exchange for other nutrients, vitamins, and protection [1, 4]. An example of this is Cyanobacteria that live inside their protist hosts (e.g., diatoms, dinoflagellates, or radiolarians) and provide photosynthesis products [65, 66] or nitrogen through nitrogen fixation [67–69], in exchange for protection and/or nutrients. There were also many records of heterotrophic bacteria engaged in symbiosis with protists in PIDA, where bacteria provide their hosts with vitamins and other types of nutrients in exchange for photosynthesis products (carbohydrates), nutrients, or protection. Such types of symbiotic interactions have been demonstrated for several relationships between bacteria and microalgae [70–72]. One example is the relationship between the diatom *Pseudo-nitzschia multiseries* and *Sulfitobacter* sp. SA11. *Pseudo-nitzschia multiseries* secretes organic carbon and a sulfonated metabolite called taurine which is taken up by the *Sulfitobacter* sp., and the *Sulfitobacter* sp. bacteria respond by secreting ammonium for the diatom and then switch their preference from ammonium to nitrate, thereby promoting the growth rate of both partners involved in the symbiosis [73]. There were remarkably few studies demonstrating symbiotic relationships between two or more heterotrophic protists in the aquatic environment. The only records we found all represented the same type of symbiotic relationship between the parasite *Neoparamoeba perurans* and its kinetoplastid endosymbionts. *Neoparamoeba perurans* is a well-studied organism since it causes disease in salmon, and consequently is a threat for aquaculture [74–77].

Compared with the other three categories (symbiont–host, parasite–host, and predator–prey) the unresolved interactions were heavily dominated by bacteria–protist interactions, mainly between Alpha- or Gammaproteobacteria interacting with Ciliophora, Dinoflagellata (both Alveolates), or Diatomea (Stramenopiles). Since the functional role of the relationship between these bacteria and protist hosts is not determined, future studies should focus on elucidating these interactions as they could be important components in protist holobionts. The holobiont concept includes the host and all associated microbes (eukaryotes, prokaryotes, and viruses). There has been a growing awareness that microbes associated with larger hosts (e.g., humans) have profound effects on the host's health and development [22, 78], and this most likely applies to protist holobionts too.

Parasites are present in most phylogenetic groups and hold important roles in ecosystems where they can for instance alter both the structure and dynamics of food webs [79, 80]. Parasites are likely largely underrepresented in

studies of microbial interactions with only 18% of the entries in PIDA. This is especially prominent in the light of recent results from environmental DNA surveys, which indicate that parasites are particularly diverse and abundant in marine as well as terrestrial ecosystems [9, 45, 81, 82]. Cercozoan parasites were only described from diatoms, although several of their close relatives are known to parasitize a plethora of macroscopic hosts [83–86]. Likewise, parasites belonging to different stramenopile lineages such as Peronosporomycetes (oomycetes), Labyrinthulomycetes, and Pirsonia were also registered to mostly infect diatoms (Fig. 1d). This probably reflects that diatoms have been the subject of more scientific studies than other protist hosts, although the true diversity of parasites infecting diatoms is likely larger than what is currently known [84].

The bipartite network analyses indicated that there were some parasites in PIDA that infected many different hosts (i.e., broad host range). But in general, the majority of parasites in PIDA appeared to infect few hosts, and altogether, parasite–host interactions seemed to be slightly more dominated by specialized interactions than in symbiont–host networks. For symbionts and parasites, the observed patterns could indicate that several studies have investigated these relationships from ‘the parasite/symbiont point of view’, and consequently, well-known taxa (e.g., the parasite *Amoebophrya*) have been investigated more thoroughly leading to broader host ranges. In contrast, several other symbionts/parasites have been detected only associated with one host, pointing to many specialized ‘one-to-one’ relationships in the microbial world. It could also be speculated that this pattern results from accidental detections of parasites or symbionts. For instance, research on diatoms would from time to time observe cells that are infected with a parasite or that host a symbiont, without looking more into the host range of these parasites or symbionts, as that was not the original focus of the study. Both the *symbiont–host*, *interactor–host*, and *parasite–host* categories in PIDA are dominated by host entries (~double number of hosts compared with the number of interacting partners), and this observation could support that several of the detections of symbionts, parasites, and interactors may represent ‘accidental detections’. The search for a parasite or symbiont's host range has until recently been like searching for a needle in a haystack.

In conclusion, summarizing the data on ecological interactions involving aquatic protists and other microbes from the past ~150 years allowed us to obtain a unique overview of the known interactions and derive relevant biological hypotheses. Despite the biases and knowledge gaps we identified, PIDA can be used for multiple purposes in future studies (which is beyond the scope of this work), for example: (1) to identify the functional role of a

microbe using taxonomically annotated environmental DNA sequences, (2) to investigate whether ecological interaction hypotheses that derive from association networks [23] are supported by previous studies in PIDA, and (3) to obtain information about the host range of a particular parasite, the predators of a specific prey, or the symbionts from a given host. Last but not least, our work identifies knowledge gaps that could be the focus of future research.

**Acknowledgements** This work has been supported by the projects MicroEcoSystems (Research Council of Norway 240904) as well as INTERACTOMICS (CTM2015-69936-P, MINECO, Spain) to RL. RL was supported by a Ramón y Cajal fellowship (RYC-2013-12554, MINECO, Spain). We thank four anonymous reviewers, in addition to Javier del Campo and Rannveig M. Jacobsen, for useful suggestions and comments that helped to improve this work.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Caron DA, Alexander H, Allen AE, Archibald JM, Armbrust EV, Bachy C, et al. Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat Rev Microbiol*. 2017;15:6–20.
- Falkowski P. The power of plankton. *Nature*. 2012;483:S17–20.
- Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. *Science*. 2008;320:1034–9.
- Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science*. 2015;347:1257594.
- Margulis L, Fester R. Symbiosis as a source of evolutionary innovation: speciation and morphogenesis. Cambridge, MA: MIT Press; 1991.
- Lopez-Garcia P, Eme L, Moreira D. Symbiosis in eukaryotic evolution. *J Theor Biol*. 2017;434:20–33.
- Archibald JM. Endosymbiosis and eukaryotic cell evolution. *Curr Biol*. 2015;25:R911–921.
- Cavalier-Smith T. Symbiogenesis: mechanisms, evolutionary consequences, and systematic implications. *Annu Rev Ecol Evol Syst*. 2013;44(44):145–72.
- Mahé F, de Vargas C, Bass D, Czech L, Stamatakis A, Lara E, et al. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat Ecol Evol*. 2017;1:91.
- Biard T, Stemann L, Picheral M, Mayot N, Vandromme P, Hauss H, et al. In situ imaging reveals the biomass of giant protists in the global ocean. *Nature*. 2016;532:504–7.
- Finlay BJ, Esteban GF. Freshwater protozoa: biodiversity and ecological function. *Biodivers Conserv*. 1998;7:1163–86.
- Huxley T. Zoological notes and observations made on board H.M. S Rattlesnake. III. Upon *Thalassicola*, a new zoophyte. *Ann Mag Nat Hist Ser 2*. 1851;8:433–42.
- Brandt K. Über das Zusammenleben von Thieren und Algen. *Verh Physiol Ges*. 1881;1:524–7.
- Logares R, Haverkamp THA, Kumar S, Lanzen A, Nederbragt AJ, Quince C, et al. Environmental microbiology through the lens of high-throughput DNA sequencing: Synopsis of current platforms and bioinformatics approaches. *J Microbiol Methods*. 2012;91:106–13.
- Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc Natl Acad Sci USA*. 2006;103:12115–20.
- Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of nextgeneration sequencing technologies. *Nat Rev Genet*. 2016;17:333–51.
- Pedrés-Alió C, Acinas SG, Logares R, Massana R. Marine microbial diversity as seen by high throughput sequencing. In: Gasol J, Kirchman D, editors. *Microbial ecology of the oceans*. New Jersey: Wiley-Blackwell; 2018. p. 47–87.
- Spang A, Saw JH, Jorgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 2015;521:173–9.
- Faust K, Lahti L, Gonze D, de Vos WM, Raes J. Metagenomics meets time series analysis: unraveling microbial community dynamics. *Curr Opin Microbiol*. 2015;25:56–66.
- Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol*. 2012;10:538–50.
- Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, et al. Determinants of community structure in the global plankton interactome. *Science*. 2015;348:1–10.
- Layeghifard M, Hwang DM, Guttman DS. Disentangling interactions in the microbiome: a network perspective. *Trends Microbiol*. 2017;25:217–28.
- Röttgers L, Faust K. From hairballs to hypotheses-biological insights from microbial networks. *FEMS Microbiol Rev*. 2018;42:761–80.
- Koeppen N. *Amoebophrya stycholonchae* nov. gen. et sp. (corps spiral de Fol). *Zoologischer Anz*. 1894;17:417–24.
- Adl SM, Bass D, Lane CE, Lukes J, Schoch CL, Smirnov A, et al. Revisions to the classification, nomenclature, and diversity of eukaryotes. *J Eukaryot Microbiol*. 2019;66:4–119.
- Adl SM, Simpson AG, Lane CE, Lukes J, Bass D, Bowser SS, et al. The revised classification of eukaryotes. *J Eukaryot Microbiol*. 2012;59:429–93.
- Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res*. 2014;42:D643–648.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41:D590–6.
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res*. 2007;35:7188–96.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. *Nucleic Acids Res*. 2017;45:D37–42.

31. R Core Development Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2018. <http://www.R-project.org>.
32. Dormann CF, Gruber B, Fruend J. Introducing the bipartite package: analysing ecological networks. *R News*. 2008;8:8–11.
33. Dormann CF, Fruend J, Bluthgen N, Gruber B. Indices, graphs and null models: analyzing bipartite ecological networks. *Open Ecol J*. 2009;2:7–24.
34. Dormann CF. How to be a specialist? Quantifying specialisation in pollination networks. *Network Biol*. 2011;1:1–20.
35. Bluthgen N, Menzel F, Bluthgen N. Measuring specialization in species interaction networks. *BMC Ecol*. 2006;6:9.
36. Wickham H. *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag; 2016.
37. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498–504.
38. Backman T, Girke T. *systemPipeR: NGS workflow and report generation environment*. *BMC Bioinforma*. 2016;17:388.
39. Mitra A, Flynn KJ, Tillmann U, Raven JA, Caron D, Stoecker DK, et al. Defining planktonic protist functional groups on mechanisms for energy and nutrient acquisition: incorporation of diverse mixotrophic strategies. *Protist*. 2016;167:106–20.
40. Stoecker DK, Hansen PJ, Caron DA, Mitra A. Mixotrophy in the Marine Plankton. *Ann Rev Mar Sci*. 2017;9:311–35.
41. Selosse MA, Charpin M, Not F. Mixotrophy everywhere on land and in water: the *grand écart* hypothesis. *Ecol Lett*. 2017;20:246–63.
42. Flynn KJ, Mitra A, Anestis K, Anschütz AA, Calbet A, Ferreira GD, et al. Mixotrophic protists and a new paradigm for marine ecology: where does plankton research go now? *J Plankton Res*. 2019;41:375–91.
43. Tanifuji G, Cenci U, Moog D, Dean S, Nakayama T, David V, et al. Genome sequencing reveals metabolic and cellular interdependence in an amoeba-kinetoplastid symbiosis. *Sci Rep*. 2017;7:11688.
44. Fontaine C, Guimaraes PR Jr., Kefi S, Loeuille N, Memmott J, van der Putten WH, et al. The ecological and evolutionary implications of merging different types of networks. *Ecol Lett*. 2011;14:1170–81.
45. de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R, et al. Eukaryotic plankton diversity in the sunlit ocean. *Science*. 2015;348:1261605.
46. Massana R, Gobet A, Audic S, Bass D, Bittner L, Boutte C, et al. Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environ Microbiol*. 2015;17:4035–49.
47. Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R, et al. Patterns of rare and abundant marine microbial eukaryotes. *Curr Biol*. 2014;24:813–21.
48. Morard R, Garet-Delmas MJ, Mahe F, Romac S, Poulain J, Kucera M, et al. Surface ocean metabarcoding confirms limited diversity in planktonic foraminifera but reveals unknown hyper-abundant lineages. *Sci Rep*. 2018;8:2539.
49. Banerjee S, Schlaeppi K, van der Heijden MGA. Keystone taxa as drivers of microbiome structure and functioning. *Nat Rev Microbiol*. 2018;16:567–76.
50. Hopkinson BM, Dupont CL, Allen AE, Morel FM. Efficiency of the CO<sub>2</sub>-concentrating mechanism of diatoms. *Proc Natl Acad Sci USA*. 2011;108:3830–7.
51. Hallegraeff GM, Jeffrey SW. Annually recurrent diatom blooms in spring along the New-South-Wales coast of Australia. *Aust J Mar Freshw Res*. 1993;44:325–34.
52. Granéli E, Turner JT. *Ecology of harmful algae*. New York: Springer, Berlin; 2008.
53. Pernthaler J. Predation on prokaryotes in the water column and its ecological implications. *Nat Rev Microbiol* 2005;3:537–46.
54. Sherr EB, Sherr BF. Capacity of herbivorous protists to control initiation and development of mass phytoplankton blooms. *Aquat Microb Ecol*. 2009;57:253–62.
55. Del Giorgio PA, Williams PJL. *Respiration in aquatic ecosystems*. Oxford, New York: Oxford University Press; 2005.
56. Armengol L, Calbet A, Franchy G, Rodriguez-Santos A, Hernandez-Leon S. Planktonic food web structure and trophic transfer efficiency along a productivity gradient in the tropical and subtropical Atlantic Ocean. *Sci Rep*. 2019;9:2044.
57. Richards TA, Leonard G, Mahe F, Del Campo J, Romac S, Jones MD, et al. Molecular diversity and distribution of marine fungi across 130 European environmental samples. *Proc Biol Sci*. 2015;282:20152243.
58. Yi Z, Berney C, Hartikainen H, Mahamdallie S, Gardner M, Boenigk J, et al. Highthroughput sequencing of microbial eukaryotes in Lake Baikal reveals ecologically differentiated communities and novel evolutionary radiations. *FEMS Microbiol Ecol*. 2017;93:fix073.
59. Grossart HP, Van den Wyngaert S, Kagami M, Wurzbacher C, Cunliffe M, RojasJimenez K. Fungi in aquatic ecosystems. *Nat Rev Microbiol*. 2019;17:339–54.
60. Lepere C, Ostrowski M, Hartmann M, Zubkov MV, Scanlan DJ. In situ associations between marine photosynthetic picoeukaryotes and potential parasites - a role for fungi? *Environ Microbiol Rep*. 2016;8:445–51.
61. Richards TA, Jones MD, Leonard G, Bass D. Marine fungi: their ecology and molecular diversity. *Ann Rev Mar Sci*. 2012;4:495–522.
62. Lukes J, Flegontova O, Horak A. Diplonemids. *Curr Biol*. 2015;25:R702–4.
63. Sherr EB, Sherr BF. Significance of predation by protists in aquatic microbial food webs. *Antonie Van Leeuwenhoek*. 2002;81:293–308.
64. Azam F, Fenchel T, Field JG, Gray JS, Meyerreil LA, Thingstad F. The ecological role of water-column microbes in the sea. *Mar Ecol Prog Ser*. 1983;10:257–63.
65. Foster RA, Carpenter EJ, Bergman B. Unicellular cyanobionts in open ocean dinoflagellates, radiolarians, and tintinnids: ultrastructural characterization and immunolocalization of phycoerythrin and nitrogenase. *J Phycol*. 2006;42:453–63.
66. Foster RA, Collier JL, Carpenter EJ. Reverse transcription PCR amplification of cyanobacterial symbiont 16S rRNA sequences from single non-photosynthetic eukaryotic marine planktonic host cells. *J Phycol*. 2006;42:243–50.
67. Gordon N, Angel DL, Neori A, Kress N, Kimor B. Heterotrophic dinoflagellates with symbiotic cyanobacteria and nitrogen limitation in the Gulf of Aqaba. *Mar Ecol Prog Ser*. 1994;107:83–8.
68. Foster RA, Kuypers MM, Vagner T, Paerl RW, Musat N, Zehr JP. Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses. *ISME J*. 2011;5:1484–93.
69. Foster RA, O'Mullan GD. Nitrogen-fixing and nitrifying symbioses in the marine environment. In: Capone DG, Bronk DA, Mulholland MR, Carpenter EJ, editors. *Nitrogen in the marine environment*. 2nd ed. Elsevier; 2008. pp 1197–218.
70. Ramanan R, Kim BH, Cho DH, Oh HM, Kim HS. Algae-bacteria interactions: evolution, ecology and emerging applications. *Biotechnol Adv*. 2016;34:14–29.
71. Cole JJ. Interactions between bacteria and algae in aquatic ecosystems. *Annu Rev Ecol Syst*. 1982;13:291–314.
72. Cooper MB, Smith AG. Exploring mutualistic interactions between microalgae and bacteria in the omics age. *Curr Opin Plant Biol*. 2015;26:147–53.
73. Amin SA, Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature*. 2015;522:98–101.
74. Dykova I, Nowak B, Peckova H, Fiala I, Crosbie P, Dvorakova H. Phylogeny of *Neoparamoeba* strains isolated from marine fish and

- invertebrates as inferred from SSU rDNA sequences. *Dis Aquat Organ.* 2007;74:57–65.
75. Dykova I, Fiala I, Peckova H. *Neoparamoeba* spp. and their eukaryotic endosymbionts similar to *Perkinsela amoebae* (Hollande, 1980): coevolution demonstrated by SSU rRNA gene phylogenies. *Eur J Protistol.* 2008;44:269–77.
76. Young ND, Dykova I, Crosbie PB, Wolf M, Morrison RN, Bridle AR, et al. Support for the coevolution of *Neoparamoeba* and their endosymbionts, *Perkinsela amoebae*-like organisms. *Eur J Protistol.* 2014;50:509–23.
77. Caraguel CG, O’Kelly CJ, Legendre P, Frasca S Jr., Gast RJ, Despres BM, et al. Microheterogeneity and coevolution: an examination of rDNA sequence characteristics in *Neoparamoeba pemaquidensis* and its prokinetoplastid endosymbiont. *J Eukaryot Microbiol.* 2007;54:418–26.
78. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. *Nat Rev Genet.* 2012;13:260–70.
79. Lafferty KD, Allesina S, Arim M, Briggs CJ, De Leo G, Dobson AP, et al. Parasites in food webs: the ultimate missing links. *Ecol Lett.* 2008;11:533–46.
80. Amundsen PA, Lafferty KD, Knudsen R, Primicerio R, Klemtsen A, Kuris AM. Food web topology and parasites in the pelagic zone of a subarctic lake. *J Anim Ecol.* 2009;78:563572.
81. Park MG, Yih W, Coats DW. Parasites and phytoplankton, with special emphasis on dinoflagellate infections. *J Eukaryot Microbiol.* 2004;51:145–55.
82. Skovgaard A. Dirty tricks in the plankton: diversity and role of marine parasitic protists. *Acta Protozoologica.* 2014;53:51–62.
83. Hartikainen H, Ashford OS, Berney C, Okamura B, Feist SW, Baker-Austin C, et al. Lineage-specific molecular probing reveals novel diversity and ecological partitioning of haplosporidians. *ISME J.* 2014;8:177–86.
84. Scholz B, Guillou L, Marano AV, Neuhauser S, Sullivan BK, Karsten U, et al. Zoospore parasites infecting marine diatoms - A black box that needs to be opened. *Fungal Ecol.* 2016;19:59–76.
85. Ward GM, Bennett M, Bateman K, Stentiford GD, Kerr R, Feist SW, et al. A new phylogeny and environmental DNA insight into paramyxids: an increasingly important but enigmatic clade of protistan parasites of marine invertebrates. *Int J Parasitol.* 2016;46:605–19.
86. Sierra R, Canas-Duarte SJ, Burki F, Schwelm A, Fogelqvist J, Dixelius C, et al. Evolutionary origins of rhizarian parasites. *Mol Biol Evol.* 2016;33:980–3.
87. Schulz F, Eloë-Fadrosch EA, Bowers RM, Jarett J, Nielsen T, Ivanova NN, et al. Towards a balanced view of the bacterial tree of life. *Microbiome.* 2017;5:140.