Neobodonids are dominant kinetoplastids in the global ocean

Running title: Kinetoplastids in the oceans

Olga Flegontova[1,2], Pavel Flegontov[1,3], Shruti Malviya[4,5], Julie Poulain[6], Colomban de Vargas[7,8], Chris Bowler[5], Julius Lukeš[1,2,*] & Aleš Horák[1,2,*]

[1] *Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic*

[2] *Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic*

[3] *Life Science Research Centre, Faculty of Science, University of Ostrava, Ostrava, Czech Republic*

[4] *Simons Centre for the Study of Living Machines, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India*

[5] *Ecole Normale Supérieure, PSL Research University, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS UMR 8197, INSERM U1024, 46 rue d'Ulm, F-75005 Paris, France*

[6] *Genoscope, CEA, Évry, France*

[7] *Station Biologique de Roscoff, Roscoff, France.*

[8] *Sorbonne Universités, Paris, France.*

[*] Corresponding authors

Corresponding address: Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Branišovská 31, 37005 České Budějovice, Czech Republic

Tel. +420387775409, +420387775403

Fax. +420385310388

Email. ogar@paru.cas.cz

Conflict of interest: The authors declare no conflict of interest

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.14034

Summary

Kinetoplastid flagellates are comprised of basal mostly free-living bodonids and derived obligatory parasitic trypanosomatids, which belong to the best-studied protists. Due to their omnipresence in aquatic environments and soil, the bodonids are of ecological significance. Here we present the first global survey of marine kinetoplastids and compare it with the strikingly different patterns of abundance and diversity in their sister clade, the diplonemids. Based on analysis of 18S rDNA V9 ribotypes obtained from 124 sampling sites collected during the *Tara* Oceans expedition, our results show generally low to moderate abundance and diversity of planktonic kinetoplastids. Although we have identified all major kinetoplastid lineages, 98% of kinetoplastid reads are represented by neobodonids, namely specimens of the *Neobodo* and *Rhynchomonas* genera, which make up 59% and 18% of all reads, respectively. Most kinetoplastids have small cell size (0.8 – 5 µm) and tend to be more abundant in the mesopelagic as compared to the euphotic zone. Some of the most abundant operational taxonomic units have distinct geographical distributions, and three novel putatively parasitic neobodonids were identified, along with their potential hosts.

Introduction

Kinetoplastid flagellates (Kinetoplastea) belong to the phylum Euglenozoa (Adl *et al.*, 2012). Basal kinetoplastid lineages are generally called bodonids, a polyphyletic assemblage of pear-shaped bi-flagellated protists. A small group of parasitic or endosymbiont species within bodonids, represented by the *Ichthyobodo* and *Perkinsela*, belongs to the clade Prokinetoplastina. However, the bulk of bodonids described so far belong to the clade Metakinetoplastina, which also includes the crown group of trypanosomatids (Moreira *et al.*, 2004). Within Metakinetoplastina, three lineages termed Eu-, Neo- and Parabodonida are recognized (Lukeš *et al.*, 2014; Yazaki *et al.*, 2017). These bodonid lineages harbor mostly free-living bacteriovores from aquatic environments and soil, where they usually constitute a relatively minor group of uncertain ecological significance (Glaser *et al.*, 2014; Atkins *et al.*, 2000; López-García *et al.*, 2003). However, the most common bodonid genera, such as *Neobodo* and *Rhynchomonas,* are considered to belong among the most important protistan bacteriovores of both the oceanic and freshwater ecosystems (Chavez-Dozal et al. 2013; Mukherjee et al. 2015) with strong and specific response to bodonid grazing from bacterial communities (Chavez-Dozal et al. 2013; Pernthaler et al., 1997)

The trypanosomatids, which encompass a majority of known species, have adopted commensal or parasitic life strategies (Lukeš *et al.*, 2014). Due to extreme diversification and host-parasite co-evolution, it seems plausible that every terrestrial vertebrate species harbors its own *Trypanosoma* species (Hamilton *et al.*, 2007). Although the medically and veterinarily important members of the

genera *Trypanosoma* and *Leishmania* have received most attention, it is within insect hosts where most of the diversity of these terrestrial parasites seems to be hidden (Maslov *et al.*, 2013).

While local protistan richness in some oceanic ecosystems has been well documented, until recently no systematic study of eukaryotic planktonic biodiversity across the world's ocean, and across the full range of organismal sizes, was available. One of the first studies to tackle this shortcoming was based on samples collected by the *Tara* Oceans expedition, by taking advantage of a huge dataset of 18S rDNA V9 metabarcode sequences to explore the taxonomic structure, ecological roles and mutual interactions of planktonic prokaryotes and eukaryotes (Brum *et al.*, 2015; de Vargas *et al.*, 2015; Lima-Mendez *et al.*, 2015; Villar *et al.*, 2015; Sunagawa *et al.*, 2015). A number of follow-up studies further focused on particular taxonomic groups of planktonic eukaryotes, analyzing in detail the V9 ribotypes and associated data (Le Bescot *et al.*, 2016; Malviya *et al.*, 2016; Mutsuo *et al.*, 2016; Flegontova *et al.*, 2016).

While a global analysis of kinetoplastid protists in marine habitats is lacking, there are a few reports from pelagic systems (von der Heyden and Cavalier-Smith, 2005; Scheckenbach *et al.*, 2010; Salani *et al.*, 2012), deep-sea benthos (Atkins *et al.*, 2000; López-García *et al.*, 2003; Brown and Wolfe, 2006; Sauvadet *et al.*, 2010;) and hypersaline anoxic basins (Edgcomb *et al.*, 2011). Neobodonids in general and *Rhynchomonas* in particular were reported as predominant kinetoplastids in sediment-overlaying water on abyssal plains (Scheckenbach *et al.*, 2010; Salani *et al.*, 2012). With the exception of the hypersaline anoxic niche, kinetoplastids generally seem to constitute a minor component of the plankton. However, this viewpoint has been challenged by the unexpectedly frequent appearance of these flagellates in FISH-based analyses of the mesopelagic and deeper layers (Morgan-Smith *et al.*, 2011). Moreover, free-living kinetoplastids have been identified as a dominant group of the hypolimnion of a freshwater lake ecosystem (Mukherjee *et al.*, 2015). These results suggest that bodonids may represent a major bacteriovorous component of the plankton that has so far escaped broader recognition (Mukherjee *et al.*, 2015).

We now know that heterotrophic protists constitute a much more diverse component of the plankton than formerly appreciated, significantly exceeding all photosynthetic eukaryotes in species number (de Vargas *et al.*, 2015; Worden *et al.*, 2015). The aim of the current study was therefore to investigate community structure, patterns of diversity and abundance, and possible ecological role of marine planktonic kinetoplastids, evaluated for the first time on a global scale in samples collected during the *Tara* Oceans expedition.

Materials and Methods

Dataset composition

We worked with the eukaryotic small subunit ribosomal RNA (18S rDNA) metabarcoding dataset obtained in the frame of the *Tara* Oceans expedition (Karsenti *et al.*, 2011; de Vargas *et al.*, 2015). The dataset included DNA sequencing reads of the V9 region of the 18S rRNA gene clustered into ribotypes. Planktonic DNA samples were collected at 124 stations worldwide (Suppl. Table 1) in eight oceanographic provinces, namely the Mediterranean Sea (MS), Red Sea (RS), Indian Ocean (IO), South Atlantic Ocean (SAO), Southern Ocean (SO), South Pacific Ocean (SPO), North Pacific Ocean (NPO), and North Atlantic Ocean (NAO). Up to three depth zones were sampled per station: surface (SRF, 5-25 m), deep chlorophyll maximum (DCM, 17-185 m), and mesopelagic zone (MES, 347-852 m). At few stations, oxygen-depleted waters were sampled (OMZ, 268-595 m). The SRF and DCM zones included up to four size fractions (Suppl. Table. 1): 0.8-5 µm (piconano-plankton), 5-20 µm (nano-plankton), 20-180 µm (micro-plankton), and 180-2,000 µm (meso-plankton), plus some additional size fractions in a few samples (>0.8 µm, 0.8-20 µm). The OMZ included three size fractions: 0.8-5 µm, 5-20 µm, and 20-180 µm. Mesopelagic samples included two size fractions: 0.8-3 µm and >3 µm. DNA was extracted from all samples, and the hyper-variable V9 region of the nuclear 18S rDNA was PCR-amplified (Amaral-Zettler *et al.*, 2009). Identical reads were merged into ribotypes, which received taxonomic assignments through annotation against an expert-curated V9 reference database (for details, see de Vargas *et al.*, 2015) derived from the PR2 database (Guillou *et al.*, 2013). Subsequently ribotypes with abundance less than 3 reads were removed in order to avoid potential biases associated with sequencing errors, following the approach used by de Vargas *et al*. (2015). The ribotypes were clustered into OTUs using the linkage clustering 'Swarm' approach relying on a local clustering threshold of a single nucleotide substitution or insertion/deletion (Mahé *et al.*, 2014, 2015). Network-based clustering methods, including Swarm, were reported to approximate true limits of sequence diversity and correspond better to taxonomic boundaries, as compared to centroid-based clustering with a global similarity threshold (Forster *et al.*, 2016; Mahe *et al., 2015*). From the resulting global dataset we extracted OTUs assigned to the Kinetoplastea phylum and refined clade assignments using an in-house 18S rRNA reference database for kinetoplastids and the ggsearch36 software.

In-house 18S rRNA reference database for kinetoplastids

For the phylogenetic analysis of nearly full-length 18S rRNA sequences we used the following approach. First, a core set of kinetoplastid 18S rRNA sequences taken from the PR2 database (Guillou *et al.*, 2013) was used as the initial query for an iterative search in the GenBank database, implemented

in the BlastCircle v.0.3 script (eukref.org). Second, the output sequences were clustered with the 97% identity threshold using USEARCH (Edgar, 2010), resulting in a set of 'seed' rRNAs. Third, MAFFT v.7.245 (Katoh and Standley, 2013) with the '--auto' option and trimAl v.1.2 (Capella-Gutiérrez *et al.*, 2009) with the '-gt 0.3' and '-st 0.001' options were used to make and prune sequence alignments, including a distant eukaryotic outgroup. Fourth, FastTree v.2.1.8 (Price *et al.*, 2010) was used to make a preliminary maximum likelihood tree. Seeds and corresponding sequence clusters falling outside of Euglenozoa were removed subsequently, and the clustering, alignment, and tree building steps were repeated a number of times until no sequences falling between the outgroup and the Euglenozoa clade were left. The final alignment was used to build a maximum likelihood tree with RAxML v.8.2.3 (Stamatakis, 2014) with the following options: phylogenetic model GTR+CAT+I; 25 rate categories; model optimization precision, 0.001; a random starting tree; 1,000 random bootstrap replicates and 200 iterations of the maximum likelihood algorithm.

Global OTU distribution analysis

The final dataset, a matrix of V9 read counts for OTUs vs. samples (Suppl. Table 1), was used for calculating the following statistics in separate samples or their combinations based on depth zones and size fraction: i/ relative abundance, i.e. the percentage of kinetoplastid V9 reads among eukaryotic V9 reads; ii/ richness, i.e. the number of kinetoplastid OTUs; iii/ Shannon diversity index, iv/ evenness. The one-way analysis of variance (ANOVA) combined with Tukey's honest significance test was used to compare the distributions across depth zones, size fractions, oceanic provinces, or three latitude zones: i/ tropical, 24°N-24°S; ii/ temperate 25-44°N and 25-44°S; iii/ Antarctic 44-65°S. Multi-way ANOVA assessed the influence of four variables listed above and their pairwise interactions on the abundance and diversity statistics. Plotting of various statistics on the world map was performed using an open source software QGIS v.2.8 (http://qgis.org/en/site/) with open-source maps. Compositional similarity between stations and oceanic provinces were computed based on Hellinger-transformed abundance matrix and incidence matrix using Bray-Curtis and Jaccard indices respectively, as a measure of β-diversity. The agglomerative method used for hierarchical cluster analysis was the Ward clustering.

Results

Abundance of kinetoplastids across depth zones and geographical regions

From the global meta-barcoding dataset from *Tara* Oceans we extracted 1 570 025 kinetoplastid reads belonging to 8 207 ribotypes clustered into 512 Operational Taxonomic Units (OTUs; see sample information and read counts for each OTU in Suppl. Table 1). Their diversity in the global dataset was

comparable to that of rhodophytes or cryptophytes, was about 15% of the richness of another major terrestrial parasitic clade, apicomplexans but only about 1% compared to their sister lineage, diplonemids. The trypanosomatids, a lineage of terrestrial parasites, were essentially missing from our samples (Table 1). The same was true for other parasitic and/or endosymbiotic lineages (*Perkinsela* and *Ichthyobodo*). The vast majority of reads (99.8%), on the other hand, belonged to free living eu- (1.3%) and especially neobodonids (98.4%; of which 59.1% belong to the *Neobodo* and 18.2% to the *Rhynchomonas*). The distribution of richness among clades correlated well with the abundance (Pearson's $r = 0.946$), although the rare lineages listed above represented a much larger fraction of kinetoplastid richness compared to their abundances (Table 1).

Rarefaction curves revealed that kinetoplastid diversity was saturated in the whole dataset, as it was for the four most diverse and abundant groups (Neobodonida, *Neobodo*, *Rhynchomonas*, and unknown Neobodonida), with rarefaction curve slopes ranging from 1×10^{-7} to 5×10^{-7} (Fig. 1). These values are comparable with the saturation of major planktonic eukaryotic groups, such as Metazoa, Dinozoa and Rhizaria. Kinetoplastid diversity was also approximately 10 times more saturated than that of diplonemids, their sister group, which were previously found to be highly diverse (Flegontova *et al.*, 2016).

Next, using one-way ANOVA, we explored the distribution of neobodonids as a whole, and of the 14 most abundant kinetoplastid OTUs (one eu- and 13 neobodonids) across three depth zones, six size fractions, three latitude zones, and eight oceanic provinces (Fig. 2). Globally, kinetoplastids represent only a small fraction of planktonic eukaryotes. Their relative abundance, i.e., the number of kinetoplastid reads divided by the number of total eukaryotic reads, averaged 0.2% per sample (ranging from 0% to 14.8%). Because neobodonids accounted for an overwhelming majority of marine kinetoplastids, the abundance of these groups mirrored the global patterns, in being significantly more abundant in the deeper mesopelagic (MES) zone below 200 metres as compared to the sunlit surface (SRF) and deep chlorophyll maximum (DCM) zones (see Methods for details). Their abundance was furthermore maximal in the smallest piconano-plankton size fractions (0.8-5 µm or 0.8-3 µm) in all zones. Thus, marine kinetoplastids are even smaller than the related diplonemids, which were almost equally abundant in the 0.8-5 µm and 5-20 µm fractions (Flegontova *et al.*, 2016). No statistically significant geographical patterns in abundance could be detected for the neobodonid group as a whole (Fig. 2).

When abundant OTUs were examined individually, we observed that the most abundant kinetoplastid OTU (belonging to *Neobodo*) as well as one eubodonid OTU were preferentially found in mesopelagic samples with low oxygen concentration (the oxygen minimum zone, OMZ). The single eubodonid taxon was found in the 5-20 µm fraction, but a majority of the abundant OTUs occurred

mostly in piconano-plankton (below 5 µm), suggesting that the organisms are likely free-living. However, three OTUs among unclassified neobodonids showed significant enrichment in the largest meso-plankton fraction (180-2000 µm). From this size distribution we infer that these are probably novel and abundant parasitic taxa (see below for details).

We also analyzed geographic distributions of the 14 most abundant OTUs (Fig. 2). Their distribution across *Tara* Oceans stations is shown in Suppl. Fig. 1, with selected maps presented in Fig. 3. *Neobodo* OTU #2753 had a peculiar distribution (Fig. 3A), being found both in the Mediterranean Sea and in the Drake Passage (supported by one-way and multi-way ANOVA), whila another *Neobodo* (OTU #3211) occurred almost exclusively in the North Atlantic Ocean (Fig. 3B). The most abundant OTU, *Neobodo* OTU #324 accounting for 36% of all kinetoplastid reads, occurred predominantly in the Indian and South Pacific Oceans (Suppl. Fig. 1). A multi-way ANOVA analysis supports the conclusion that the abundance of this OTU depends on two variables: size fraction and oceanic province (Fig. 4). *Neobodo* OTU #1514 occurred mostly in the tropical latitudes (Suppl. Fig. 1), but the effect is statistically significant in the mesopelagic zone only (Fig. 2). Thus, four of the five most abundant *Neobodo* OTUs displayed distinct biogeographies. *Rhynchomonas* OTU #678 was prevalent at tropical latitudes (supported by one-way ANOVA, Suppl. Fig. 1), and *Rhynchomonas* OTU #3853 had a rather narrow geographic distribution (Fig. 3C), occurring almost exclusively at tropical latitudes of the Pacific Ocean and in the North Atlantic Ocean (supported by multi- and one-way ANOVA, see Figs. 2 and 4). Among six OTUs belonging to unknown neobodonids, only two demonstrated a geographic pattern (Suppl. Fig. 1): a putatively parasitic OTU #3742 was dominant in tropical regions, whereas OTU #3677 was more widely distributed in the North Atlantic, tropical Pacific and Indian Oceans (both cases are supported by one-way ANOVA). The only eubodonid OTU in our dataset, OTU #2803, was somewhat prevalent in the Mediterranean Sea although this effect was not statistically significant (Suppl. Fig. 1).

Diversity

We then examined the effect of depth, size fraction, latitudinal gradients and oceanic provinces on kinetoplastid diversity (Suppl. Fig. 2). We analyzed all kinetoplastids, neobodonids only, or the most abundant neobodonid sub-clades (neobodonids account for about 70% of kinetoplastid OTUs, see Table 1). Diversity of all kinetoplastids and neobodonids exhibited very similar patterns (Suppl. Fig. 2), and generally followed the same trends as their relative abundance, peaking in the MES zone and in the piconano-plankton size fraction. The same patterns were observed for all three major neobodonid sub-groups, apart from *Rhynchomonas*, whose richness was not significantly stratified by depth. However, when considering the number of OTUs unique to certain depth zones and size

fractions, we observed that the surface zone had by far the largest number of unique OTUs: 37% of all kinetoplastid OTUs were unique to this zone (Fig. 5A). In contrast, just 5% of OTUs were unique to the mesopelagic zone, even though average richness per station was much higher in this zone (Fig. 2). An explanation for this apparent paradox could be found by analysis of occupancy, because we found that the vast majority of surface-specific OTUs occurred in just a few stations (Fig. 5).

On the other hand, the distribution of unique OTUs across size fractions was better correlated with richness: 21% of kinetoplastid OTUs were unique to the piconano-plankton fraction (Fig. 5B), and the same fraction demonstrated the highest richness (Fig. 2). Notably, about 3% of OTUs were unique to the micro-plankton fraction, and the same percentage was observed for the meso-plankton. These OTUs may represent parasitic species of low abundance.

Evenness and richness followed different trends: evenness peaked in DCM samples for all kinetoplastids except *Neobodo*, for which it was maximal in the MES zone (Fig. 4). For all kinetoplastids except *Rhynchomonas*, evenness peaked in the nano- and micro-plankton size fractions, while it was the piconano- and nano-plankton in the case of *Rhynchomonas*. The richness of kinetoplastids, neobodonids, *Rhynchomonas*, and unknown neobodonids was significantly higher in tropical regions (Fig. 4), in particular in the South Pacific Ocean, although this effect may be due to a higher proportion of tropical samples from that region (72%). However, *Rhynchomonas* richness was the highest in the North Pacific and North Atlantic Oceans. Evenness demonstrated no statistically significant differences across latitudes or provinces (Fig. 2).

Ecological interactions

While the majority of kinetoplastid reads were found in the pico-nano size fraction, suggesting their small cell size, our results also showed a part of kinetoplastid diversity that was associated with larger fractions. These are putative candidates for parasitic/symbiotic lifestyles. An analysis of abundant OTUs revealed such associations for three neobodonid OTUs. The only parasitic neobodonid described so far is *Azumiobodo* (Hirose *et al.*, 2012; Kumagai *et al.*, 2013; Yazaki *et al.*, 2017), which was poorly represented within our samples (about 7,000 reads in total). Significantly more abundant were the novel parasite candidates (with 15 000, 29 000 and 45 000 reads, respectively). By analyzing a global interaction network for planktonic OTUs within the photic zone (Lima-Mendez *et al.*, 2015), we found only the latter putatively parasitic OTU (OTU #2083) amongst the interactions meeting the inclusion criteria of this former work. The other two OTUs were lacking in the interactome since they were mainly found in the mesopelagic zone (Fig. 2). The geographic distribution of OTU #2083 is shown in Fig. 3D. This OTU demonstrated putative interactions with 26 taxa, mostly bacteria and alveolates, but possibly the most interesting finding was a co-presence with a planktonic

appendicularian species *Megalocercus huxleyi* occurring in the meso-plankton fraction, where OTU #2083 was also relatively abundant (Fig. 2). Notably, *Azumiobodo* parasitizes on ascidians, a closely related group of benthic animals (Hirose *et al.*, 2012; Kumagai *et al.*, 2013), and another neobodonid, *Cruzella marina*, is a commensal in the ascidian intestine (Frolov and Malysheva, 2002).

To find hosts of the other two putatively parasitic OTUs, we applied a simple approach: calculated Pearson's correlation coefficients for the 14 most abundant kinetoplastid OTUs (or kinetoplastid sub-clades) vs. 456 abundant metazoan OTUs. We used absolute abundance values (read counts) and considered metazoans represented by more than 10 000 reads. The best correlation among all kinetoplastid OTUs tested was between the putatively parasitic OTU #4802 and a copepod OTU belonging to the Calanoida group, $r = 1$, p-value = 0. Thus, we found possible hosts for two out of three putatively parasitic OTUs: an appendicularian and a copepod.

We then analyzed possible interactions of other kinetoplastids found in euphotic zone (SRF and DCM) samples. The interactome (http://www.raeslab.org/companion/ocean-interactome.html) contains only 12 kinetoplastid ribotypes (belonging to 12 OTUs) meeting the stringent inclusion criteria (Lima-Mendez et al., 2015). OTUs outside of neobodonids 'interact' with none or few (less than 10) other ribotypes, therefore no conclusive interpretation of the interactome is possible for these groups. For neobodonids, 208 positive (co-occurrence of a kinetoplastid ribotype with another ribotype) and 27 negative putative interactions (mutual exclusions) that cannot be explained by environmental factors affecting both interacting organisms (Suppl. Table 2) were found. Co-presence with bacteria and archaea, their main food source, was detected (42 interactions), as well as 34 co-occurrences with other bacteriovorous protists, such as ciliates, choanoflagellates, foraminiferans, radiolarians, and marine stramenopiles (MAST). The largest fraction of putative positive interactions (78 instances) involves various groups of Syndiniales (MALV) and other dinoflagellates. Most instances of mutual exclusion include crustaceans, cnidarians, molluscs, and ascidians (16 of 27 negative interactions), yet metazoans also participate in putative positive interactions (18 instances). We also expected to see a strong positive correlation between the intracellular symbiont *Perkinsela* and its host amoeba, however this correlation was not found in the interactome (Lima-Mendez *et al.*, 2015) because the abundance of *Perkinsela* was too low. However, using the simpler approach a very strong correlation was revealed between *Perkinsela* (all OTUs combined) and the most abundant *Paramoeba* OTU annotated as *Paramoeba branchifila*: $r = 0.98$, p-value = 0.

Discussion

There are generally two opposing views on the modes and limits of dispersal of protists. The first one champions the idea of a cosmopolitan, ubiquitous distribution, the main driving force being their short

generation times, a high rate of dispersal, large population sizes and ability to form resistant cysts (Finlay and Fenchel, 2004; Boenigk *et al.*, 2012). In this scenario, biogeographies are shaped mostly by environmental selective pressures, and similar environments across the globe are expected to support similar communities. This postulate of "everything is everywhere, but the environment selects" is challenged by an alternative view that assumes limited dispersal rates for at least some protists and finds their level of endemism comparably high with respect to other eukaryotes (Foissner, 2006). A detailed analysis of two closely related groups of excavates, kinetoplastids and diplonemids, in a truly global dataset composed of samples covering all oceanic provinces, allows interrogation of the above contrasting scenarios from a new perspective.

Both diplonemids and kinetoplastids are heterotrophic protists of similar size (mostly present in the piconano fraction) that are more abundant in the MES zone as compared to the euphotic zone, a pattern that is characteristic for a number of marine heterotrophs (Pernice *et al.*, 2015; Worden *et al.*, 2015). Although very little is known about the lifestyle of the former group, based on their distribution across size fractions, only a small portion of the described OTUs are likely to be parasites (Gawryluk *et al.*, 2016; Flegontova *et al.*, 2016). The same seems to be the case for marine kinetoplastids, because the confirmed parasitic groups constitute a mere ~0.5% of reads: all Trypanosomatida, *Ichthyobodo* (Prokinetoplastina), *Azumiobodo* (Neobodonida), *Trypanoplasma* and *Cryptobia* (Parabodonida). Three potentially parasitic OTUs found in this study account for 5.7% of reads. With the available data, one would therefore assume that most oceanic diplonemids and kinetoplastids are free-living and may thus have rather similar ecological roles. Yet unexpectedly, the patterns of their abundance and diversity are dramatically different.

Diplonemids emerged recently as the most diverse and 6th most abundant eukaryotic taxon in the global plankton (Gawryluk *et al.*, 2016; Flegontova *et al.*, 2016; David and Archibald, 2016). In contrast, in previous reports kinetoplastids constituted just a minor component of pelagic communities, being significantly more abundant in (abyssal) benthic communities (Salani *et al.*, 2012; Sauvadet *et al.*, 2010; Atkins *et al.*, 2000; Brown and Wolfe, 2006; Scheckenbach *et al.*, 2010). The only marine habitat with a significant content of kinetoplastids reported so far are hypersaline anoxic basins (Edgcomb *et al.*, 2011). From the global set of 124 examined sampling sites, we have confirmed that, with few exceptions, kinetoplastids indeed constitute a small component of the marine plankton, since their average relative abundance among eukaryotes was only 0.2%. Our results do not confirm the increased presence of kinetoplastids in oxygen-depleted habitats, although two of the 14 most abundant kinetoplastid OTUs were significantly more abundant at OMZ sites compared to other zones (Fig. 2). Because only nine OMZ samples (7% of sampling sites) were available in our global dataset, we have refrained from their more detailed analysis and included them among the MES samples,

where they fit based on the sampling depth.

It is known that the universal primers targeting the V9 or V4 hyper-variable 18S rRNA regions are not truly universal, i.e. are much more efficient for some eukaryotic clades as compared to others (Amaral-Zettler *et al.*, 2009; Hong *et al.*, 2009; Edgcomb *et al.*, 2011). For instance, the V4 primers are known to work poorly for excavates and foraminiferans (Pawlowski *et al.*, 2011; Pernice *et al.*, 2016), important protist groups in the marine plankton (de Vargas *et al.*, 2015). None of the two widely used barcodes is perfect, as demonstrated by Giner *et al.* (2016): a cDNA-based analysis of V4 metabarcodes produced relative abundance values closer to those estimated by FISH in five planktonic samples, while V9 performed better in four samples. The results were influenced by the community composition. The V4 region, in contrast to V9, is known to be highly variable in length across major eukaryotic clades (Pawlowski et al. 2011). Thus, not only differential primer specificity, but also amplicon length variability might make PCR less efficient for certain clades. This problem was highlighted in a study targeting the V4 region in bathypelagic samples (Pernice et al. 2016): the abundance of excavates (most-likely belonging to the diplonemid and/or kinetoplastid clade) assessed by metagenome-derived tags ("mitags") was 11%, but only 1% as estimated using the V4 tags.

We have therefore attempted to recover kinetoplastid signal from the *Tara* Oceans metagenomic data by mapping metagenomic reads from several samples with highest abundance of kinetoplastids on the exhaustive dataset of 344 18S sequences extracted from the public databases. However, no significant hits were retrieved, possibly because the 0.22 – 3 um origin of metagenomic dataset and resulting overabundance of prokaryotic reads. In summary, although the V4 region allows higher taxonomic resolution and contains more phylogenetic information, the shorter V9 region (Amaral-Zettler et al. 2009) might recover less biased community composition (Pawlowski et al. 2011). Given the known problems of V4 barcoding in excavates, the V9 region was the barcode of choice in our study.

However, low global kinetoplastid counts may still be an underestimation of the reality. It was recently demonstrated in freshwater habitats that in metabarcoding studies using universal primers kinetoplastids went largely undetected, yet were significantly more abundant when a specific set of oligonucleotides and/or *in situ* hybridization was employed (Mukherjee *et al.*, 2015). In fact, they dominated the eukaryotic communities in the latter case. Indeed, these protists were reported as being highly abundant in the Atlantic Ocean using kinetoplastid-specific FISH probes (Morgan-Smith *et al.*, 2011). Our results based on the V9 region of 18S rRNA are in agreement with other environmental clone-based studies (mentioned above), which using other sets of oligonucleotides showed the relative low abundance of kinetoplastids in the global plankton. However, until comparative studies are performed, we cannot exclude that, due to their divergent 18S rRNA sequences, they remain heavily

underestimated in the clone libraries.

The bulk of kinetoplastid abundance and diversity in the plankton falls into the neobodonid clade (~98% of all reads), while the endosymbiotic *Perkinsela* of the Prokinetoplastina clade is diverse (~8% of all kinetoplastid OTUs) but very rare in the plankton (only 0.2% of all kinetoplastid reads). Neobodonids are also morphologically the most diverse group among four bodonid clades (Lukeš *et al.*, 2014). It is worth noting that neobodonids are the main kinetoplastid clade in the soil (Ekelund *et al.*, 2001; Glaser *et al.*, 2014). Representatives of the genus *Neobodo* were dominant kinetoplastids in most stations sampled across all depth zones, including samples from the oxygen-depleted waters. The high abundance and global distribution of *Neobodo* species in our dataset was not unexpected, as they (and namely representatives of the *N. designis* complex) are known to be widely present in both marine and freshwater habitats (von der Heyden *et al.*, 2004; Lee and Patterson, 2002; Lee and Patterson, 1998; Scheckenbach *et al.*, 2006). Together with another neobodonid, *Rhynchomonas* spp., they were also virtually the only kinetoplastids for which any significant interactions could be retrieved from the global interactome of Lima-Mendez *et al*. (2015). Apart from the expected co-occurrence with their supposed bacterial prey and other bacterivorous protists, such as ciliates, choanoflagellates and MAST, we have found a surprisingly high number of interactions with various species of syndinians (MALV). These are typically parasites of various planktonic organisms. Although neobodonids were mostly found in samples from size fractions smaller than 20 μm and are thus possibly of very small cell size, we cannot exclude the possibility that neobodonids are target hosts of these parasitic syndinians. Alternatively, neobodonids and syndinians may co-infect the same hosts.

Kinetoplastids are represented by hundreds of OTUs, yet just 14 abundant OTUs accounted for 93% of all reads. A total of 13 of the hyper-abundant OTUs are assigned to neobodonids and one to eubodonids. The pattern of relatively few dominant and globally distributed OTUs is very similar to that observed for diplonemids which, on the other hand, have diversified into tens of thousands of OTUs (Lukeš *et al.*, 2015; Flegontova *et al.*, 2016; David and Archibald, 2016). Distribution of several abundant neobodonid OTUs shows significant geographic signature, which may be caused by their physiology and/or association with specific prey. Unfortunately, we could not find any apparent interaction with other organisms explaining the aforementioned biogeographic signatures. This may be caused by the fact that while neobodonids were mostly found in the deeper layers, the *in-silico* interactome (Lima-Mendez *et al.*, 2015) is available only for samples from the euphotic zone, and thus does not include the majority of the kinetoplastid data.

Two other kinetoplastid groups are worth mentioning. One is the obligatory parasitic

trypanosomatids, which constitute an absolute majority of the terrestrial diversity of kinetoplastids (Maslov *et al.*, 2013) and are likely one of the most diverse parasitic protists, yet are extremely rare in marine samples. This is not surprising because marine trypanosomatids are blood parasites mainly of fish (Woo, 2003; Lom and Dyková, 1992), a segment of marine life not specifically targeted by *Tara* Oceans and thus missing from out dataset. The second case of Prokinetoplastina is more interesting. These early-branching kinetoplastids are represented by *Ichthyobodo* spp., ectoparasites of freshwater and marine fish, and *Perkinsela* spp., endosymbionts of amoebae of the genus *Paramoeba* (Dyková *et al.*, 2003; Simpson *et al.*, 2006; Tanifuji *et al.*, 2011; Feehan *et al.*, 2013), with the latter representing a vast majority of reads assigned to this clade, yet still only 0.2% of all kinetoplastids. *Perkinsela* (but not its host amoebae) fell below the abundance threshold set in the global interactome, but read counts for *Perkinsela* (all OTUs) and *Paramoeba* (all OTUs) are strongly correlated in our dataset (Pearson's $r = 0.93$), while any other kinetoplastid clade or any of the 14 highly abundant OTUs show no correlation with *Paramoeba* (|Pearson's r| up to 0.11). These amoebae are ectoparasites of fish gills (Dyková *et al.*, 2003), and thus are expected to be missing or very rare in our dataset due to the sampling strategy targeting only microbes and the smallest metazoans. We suppose that *Paramoeba* reads we analyzed might be derived from free-floating cystic stages.

All in all, kinetoplastids follow a similar pattern as diplonemids: both groups show higher relative abundance in the MES zone and are dominated by just a handful of very abundant cosmopolitan OTUs. However, regarding their diversity there are significant differences. While diplonemids have undergone extreme (and likely recent) speciation into tens of thousands of OTUs of (mostly) low abundance, the same process appears not to have taken place in marine kinetoplastids. While the two major and opposing views on the distribution of protists, the "everything is everywhere" versus the "endemicity rules" theories fail to explain such a discrepancy, we should refrain from speculations until we learn more about the lifestyles of diplonemids. Notwithstanding, the sample richness and depth of sequencing presented here provide the first global and comprehensive insights into the qualitative and quantitative composition of kinetoplastids in the world's ocean.

Accession numbers

The project number for the sequences reported in this paper is EBI: XXXXXXXX.

Author contribution

AH, PF and JL designed the study; JP, CDV and CB provided the data; OF, PF, SM, and AH performed the data analysis; and AH, PF, OF, JL, and CB wrote the manuscript.

Acknowledgements

This work was supported by the ERC CZ LL1601 (to J.L.), and the Czech Grant Agency projects Nos. 15-17643S (to A.H.) and 14-23986S (to J.L.). C.B. acknowledges funding from the ERC Advanced Award ‘‘Diatomite’’, the Louis D Foundation, and the French Government ‘‘Investissements d’Avenir’’ programmes MEMO LIFE (ANR-10-LABX-54), PSL* Research University (ANR-1253 11-IDEX-0001-02). CB also thanks the Radcliffe Institute of Advanced Study at Harvard University for a scholars fellowship during the 2016-2017 academic year. We thank the commitment of the following people and sponsors: CNRS (in particular Groupement de Recherche GDR3280), European Molecular Biology Laboratory (EMBL), Genoscope/CEAthe French Government 'Investissements d'Avenir' programmes OCEANOMICS (ANR-11-BTBR-0008) and FRANCE GENOMIQUE (ANR-10-INBS-09-08), Agence Nationale de la Recherche, and European Union FP7 (MicroB3/No.287589), We also thank the support and commitment of agnès b. and Etienne Bourgois, the Veolia Environment Foundation, Region Bretagne, Lorient Agglomeration, World Courier, Illumina, the Eléctricité de France (EDF) Foundation, Fondation pour la recherche sur la biodiversité (FRB) , the Foundation Prince Albert II de Monaco, the Tara Foundation, its schooner and teams. We thank MERCATOR-CORIOLIS and ACRI-ST for providing daily satellite data during the expedition. We are also grateful to the French Ministry of Foreign Affairs for supporting the expedition and to the countries who graciously granted sampling permissions. *Tara* Oceans would not exist without continuous support from 23 institutes (http:// http://oceans.taraexpeditions.org/en/m/science/labs-involved/). 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. This article is contribution number ZZZ of *Tara* Oceans.

Conflict of interest: The authors declare no conflict of interest

References

Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, *et al.* (2012). The revised classification of eukaryotes. *J Eukaryot Microbiol* **59**: 429–493.

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal

RNA Genes. *PLoS One* **4**. doi: 10.1371/journal.pone.0006372.

Atkins MS, Teske AP, Anderson OR. (2000). A survey of flagellate diversity at four deep-sea hydrothermal vents in the Eastern Pacific Ocean using structural and molecular approaches. *J Eukaryot Microbiol* **47**: 400–11.

Le Bescot N, Mahé F, Audic S, Dimier C, Garet MJ, Poulain J, *et al.* (2016). Global patterns of pelagic dinoflagellate diversity across protist size classes unveiled by metabarcoding. *Environ Microbiol* **18**: 609–26.

Boenigk J, Ereshefsky M, Hoef-Emden K, Mallet J, Bass D. (2012). Concepts in protistology: Species definitions and boundaries. *Eur J Protistol* **48**: 96–102.

Brown PB, Wolfe G V. (2006). Protist genetic diversity in the acidic hydrothermal environments of Lassen Volcanic National Park, USA. *J Eukaryot Microbiol* **53**: 420–31.

Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, *et al.* (2015). Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science* **348**: 1261498.

Calduch-Giner JA, Sitjà-Bobadilla A, Pérez-Sánchez J. (2016). Gene expression profiling reveals functional specialization along the intestinal tract of a carnivorous teleostean fish (*Dicentrarchus labrax*). *Front Physiol* **7**: 359.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Chavez-Dozal A, Gorman C, Erken M, Steinberg PD, McDougald D, Nishiguchi MK (2013). From the laboratory into the field: Testing defense mechanisms of bacterial biofilms against protozoan grazing. *Appl Environ Microbiol* **79:** 553–558.

David V, Archibald JM. (2016). Evolution: Plumbing the depths of diplonemid diversity. *Curr Biol* **26**: R1290–R1292.

Dyková I, Fiala I, Lom J, Lukeš J. (2003). *Perkinsiella* amoebae-like endosymbionts of *Neoparamoeba* spp., relatives of the kinetoplastid *Ichthyobodo*. *Eur J Protistol* **39**: 37–52.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–1.

Edgcomb V, Orsi W, Bunge J, Jeon S, Christen R, Leslin C, *et al.* (2011). Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J* **5**: 1344–1356.

Ekelund F, Rønn R, Griffiths BS. (2001). Quantitative estimation of flagellate community structure and diversity in soil samples. *Protist* **152**: 301–314.

Feehan CJ, Johnson-Mackinnon J, Scheibling RE, Lauzon-Guay J-S, Simpson AGB. (2013). Validating the identity of *Paramoeba invadens*, the causative agent of recurrent mass mortality of sea urchins in Nova Scotia, Canada. *Dis Aquat Organ* **103**: 209–27.

Finlay BJ, Fenchel T. (2004). Cosmopolitan metapopulations of free-living microbial eukaryotes. *Protist* **155**: 237–44.

Flegontova O, Flegontov P, Malviya S, Audic S, Wincker P, de Vargas C, *et al.* (2016). Extreme diversity of diplonemid eukaryotes in the Ocean. *Curr Biol* **26**: 3060–3065.

Foissner W. (2006). Biogeography and dispersal of micro-organisms: A review emphasizing protists. *Acta Protozool* **45**: 111–136.

Forster D, Dunthorn M, Mahé F, Dolan,JR, Audic S, Bass D, Bittner L, Boutte C, Christen R, Claverie JM, Decelle J, Edvardsen B, Egge E, Eikrem W, Gobet A, Kooistra WHCF, Logares R, Massana R, Montresor M, Not F, Ogata H, Pawlowski J, Pernice, MC, Romac S, Shalchian-Tabrizi K, Simon N, Richards TA, Santini S, Sarno D, Siano R, Vaulot D, Wincker P, Zingone A, de Vargas C, Stoeck, T. (2016). Benthic protists: the under-charted majority. *FEMS Microbiol Ecol* **92**: fiw120.

Frolov AO, Malysheva MN (2002). Ultrastructure of the flagellate Cruzella marina (Kinetoplastidea). *Tsitologiia* **44:** 477-84.

Gawryluk RMR, del Campo J, Okamoto N, Strassert JFH, Lukeš J, Richards TA, *et al.* (2016). Morphological identification and single-cell genomics of marine diplonemids. *Curr Biol* **26**: 3053–3059.

Glaser K, Kuppardt A, Krohn S, Heidtmann A, Harms H, Chatzinotas A. (2014). Primer pairs for the specific environmental detection and T-RFLP analysis of the ubiquitous flagellate taxa Chrysophyceae and Kinetoplastea. *J Microbiol Methods* **100**: 8–16.

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, *et al.* (2013). The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* **41**: D597-604.

Hamilton PB, Gibson WC, Stevens JR. (2007). Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. *Mol Phylogenet Evol* **44**: 15–25.

von der Heyden S, Cavalier-Smith T. (2005). Culturing and environmental DNA sequencing uncover

hidden kinetoplastid biodiversity and a major marine clade within ancestrally freshwater *Neobodo designis*. *Int J Syst Evol Microbiol* **55**: 2605–21.

von der Heyden S, Chao EE, Vickerman K, Cavalier-Smith T. (2004). Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of euglenozoa. *J Eukaryot Microbiol* **51**: 402–16.

Hirose E, Nozawa A, Kumagai A, Kitamura S. (2012). *Azumiobodo hoyamushi* gen. nov. et sp. nov. (Euglenozoa, Kinetoplastea, Neobodonida): a pathogenic kinetoplastid causing the soft tunic syndrome in ascidian aquaculture. *Dis Aquat Organ* **97**: 227–235.

Je Lee W, Patterson DJ. (1998). Diversity and geographic distribution of free-living heterotrophic flagellates – analysis by PRIMER. *Protist* **149**: 229–244.

Karsenti E, Acinas SG, Bork P, Bowler C, de Vargas C, Raes J, *et al.* (2011). A holistic approach to marine eco-systems biology. *PLoS Biol* **9** doi: 10.1371/journal.pbio.1001177.

Katoh K, Standley DM. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usabiltiy. *Mol Biol Evol* **30**: 772–780.

Kumagai A, Ito H, Sasaki R. (2013). Detection of the kinetoplastid *Azumiobodo hoyamushi*, the causative agent of soft tunic syndrome, in wild ascidians *Halocynthia roretzi*. *Dis Aquat Organ* **106**: 267–71.

Lee WJ, Patterson DJ. (2002). Abundance and biomass of heterotrophic flagellates, and factors controlling their abundance and distribution in sediments of Botany Bay. *Microb Ecol* **43**: 467–481.

Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, *et al.* (2015). Determinants of community structure in the global plankton interactome. *Science* 3**48**: 1262073_1-1262073_9.

Lom J, Dyková I. (1992). Protozoan parasites of fishes. Developments in aquaculture and fisheries science. Volume 26. Elsevier: Amsterdam.

López-García P, Philippe H, Gail F, Moreira D. (2003). Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc Natl Acad Sci U S A* **100**: 697–702.

Lukeš J, Flegontova O, Horák A. (2015). Diplonemids. *Curr Biol* **25**: R702–R704.

Lukeš J, Skalický T, Týč J, Votýpka J, Yurchenko V. (2014). Evolution of parasitism in kinetoplastid flagellates. *Mol Biochem Parasitol* **195**: 115–122.

Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. (2014). Swarm□: robust and fast clustering method for amplicon-based studies PrePrints PrePrints. *PeerJ* 1–12.

Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. (2015). Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* **3**: e1420.

Malviya S, Scalco E, Audic S, Vincent F, Veluchamy A, Poulain J, *et al.* (2016). Insights into global diatom distribution and diversity in the world's ocean. *Proc Natl Acad Sci U S A* **113**: E1516-25.

Maslov DA, Votýpka J, Yurchenko V, Lukeš J. (2013). Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. *Trends Parasitol* **29**: 43–52.

Moreira D, López-García P, Vickerman K. (2004). An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. *Int J Syst Evol Microbiol* **54**: 1861–1875.

Morgan-Smith D, Herndl G, van Aken H, Bochdansky A. (2011). Abundance of eukaryotic microbes in the deep subtropical North Atlantic. *Aquat Microb Ecol* **65**: 103–115.

Mukherjee I, Hodoki Y, Nakano S-I. (2015). Kinetoplastid flagellates overlooked by universal primers dominate in the oxygenated hypolimnion of Lake Biwa, Japan. *FEMS Microbiol Ecol* **91**. doi: 10.1093/femsec/fiv083.

Mutsuo I, Lopes dos Santos A, Gourvil P, Yoshikawa S, Kamiya M, Ohki K, *et al.* (2016). Diversity and oceanic distribution of Parmales and Bolidophyceae, a picoplankton group closely related to diatoms. *Isme J* **10**: 2419–34.

Pernice M, Giner C, Logares R. (2015). Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J* **10**: 945–958.

Pernthaler J, Posch T, Simek K, Vrba J, Amann R, Psenner R. (1997). Contrasting bacterial strategies to coexist with a flagellate predator in an experimental microbial assemblage. *Appl Environ Microbiol* **63:** 596–601.

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* **5**. doi: 10.1371/journal.pone.0009490.

Salani FS, Arndt H, Hausmann K, Nitsche F, Scheckenbach F. (2012). Analysis of the community structure of abyssal kinetoplastids revealed similar communities at larger spatial scales. *ISME J* **6**: 713–23.

Sauvadet AL, Gobet A, Guillou L. (2010). Comparative analysis between protist communities from the deep-sea pelagic ecosystem and specific deep hydrothermal habitats. *Environ Microbiol* **12**: 2946–2964.

Scheckenbach F, Hausmann K, Wylezich C, Weitere M, Arndt H. (2010). Large-scale patterns in biodiversity of microbial eukaryotes from the abyssal sea floor. *Proc Natl Acad Sci U S A* **107**: 115–120.

Scheckenbach F, Wylezich C, Mylnikov AP, Weitere M, Arndt H. (2006). Molecular comparisons of freshwater and marine isolates of the same morphospecies of heterotrophic flagellates. *Appl Environ Microbiol* **72**: 6638–43.

Simpson AGB, Stevens JR, Lukeš J. (2006). The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol* **22**: 168–174.

Stamatakis A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, *et al.* (2015). Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**: 1261359.

Tanifuji G, Kim E, Onodera NT, Gibeault R, Dlutek M, Cawthorn RJ, *et al.* (2011). Genomic characterization of *Neoparamoeba pemaquidensis* (Amoebozoa) and its kinetoplastid endosymbiont. *Eukaryot Cell* **10**: 1143–6.

de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R, *et al.* (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science (80-)* **348**: 1261605–1261605.

Villar E, Farrant GK, Follows M, Garczarek L, Speich S, Audic S, *et al.* (2015). Ocean plankton. Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science* **348**: 1261447.

Woo PTK. (2003). *Cryptobia* (*Trypanoplasma*) salmositica and salmonid cryptobiosis. *J Fish Dis* **26**: 627–646.

Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. (2015). Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science* **347**: 1257594–1257594.

Yazaki E, Ishikawa SA, Kume K, Kumagai A, Kamaishi T, Tanifuji G, Hashimoto T, Inagaki Y. (2017). Global Kinetoplastea phylogeny inferred from a large-scale multigene alignment including parasitic species for better understanding transitions from a free-living to a parasitic lifestyle. *Genes Genet Syst* **92**: 35–42.

Table legends

Table 1. Summary of kinetoplastid diversity and abundance by taxonomic groups. The taxonomic assignment of OTUs against an in-house reference database (see Methods) was performed with ggsearch 36 according to de Vargas et al. (2015).

Figure legends

Figure 1. Rarefaction curves for OTUs: OTU count vs. read number. Slopes calculated for 10 last data points are indicated in the legend on the right. Curves were constructed for the full Kinetoplastea dataset, for the neobodonid clade and for its most abundant sub-groups: *Neobodo*, *Rhynchomonas*, and unknown Neobodonida.

Figure 2. Variation in average kinetoplastid abundance across depth zones, size fractions, and geographical regions. Only most abundant kinetoplastid clades and 14 most abundant OTUs were considered. The bar plots show average relative abundance, with scale at the bottom of each column; and pairs of the minus and asterisk symbols mark significant differences according to one-way ANOVA. Because kinetoplastids were preferentially found in the smallest size fraction of 0.8-5 um and in the mesopelagic zone, geographic variables were considered not only on the whole dataset, but also separately on these subsets. Furthermore, because a different set of size fractions was taken in the mesopelagic zone, the size variability was assessed in this zone separately. The following abbreviations are used: SRF, surface zone; DCM, deep chlorophyll maximum zone; OMZ, oxygen minimum zone; MES, mesopelagic zone; MS, Mediterranean Sea; RS, Red Sea; IO, Indian Ocean; SAO, South Atlantic Ocean; SO, Southern Ocean; SPO, South Pacific Ocean; NPO, North Pacific Ocean; NAO, North Atlantic Ocean.

Figure 3. Examples of geographical distribution of abundant OTUs. Two *Neobodo* OTUs (**A, B**), one *Rhynchomonas* OTU (**C**), and the putatively parasitic OTU #2083 (**D**) are shown. Full results for these and the other abundant OTUs are presented in Suppl. Figure 1. Relative abundance is color coded, see legends in each panel.

Figure 4. Factors driving abundance and diversity of kinetoplastids. We performed a multi-way ANOVA analysis to determine which variables drive relative abundance and diversity of kinetoplastids and their most abundant sub-clades. The strongest influence we observed was size fractions affecting abundance and diversity. Abundance was also significantly affected by depth, and in case of four

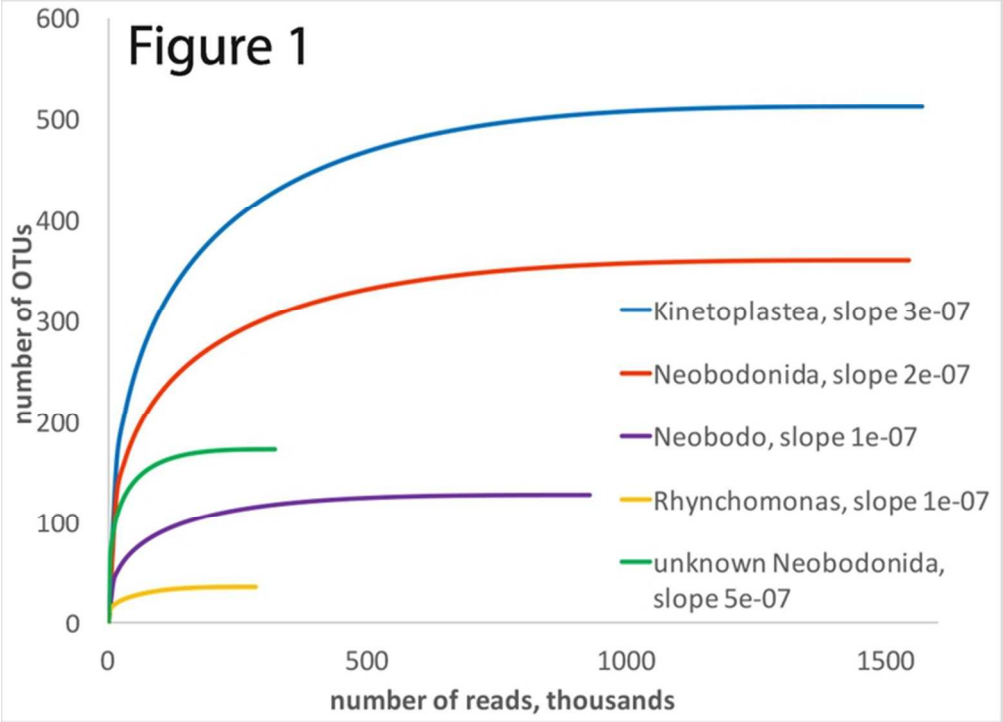
OTUs also by oceanic provinces or latitude regions. On the other hand, diversity statistics were significantly affected by all four variables. *P*-values are coded as follows: full black, *p*-values < 0.001; chequered, *p*-values from 0.001 to 0.01; horizontal stripes, *p*-values from 0.01 to 0.05.

Figure 5. Venn diagrams showing OTU distribution across depth zones (A) and size fractions (B). The following abbreviations are used: SRF, surface zone; DCM, deep chlorophyll maximum; MES, mesopelagic zone; OMZ, oxygen minimum zone.

Figure 6. Analysis of cosmopolitan and rare OTUs. Occupancy values, i.e., the number of stations where an OTU was found, are plotted on the x axis, and average station evenness for these stations is plotted on the y axis. Bubble size represents a read count for a given OTU, and OTUs unique to one depth zone (**A**) or taxonomic group (**B**) are color-coded according to the legend.

Table 1. Summary of kinetoplastid diversity and abundance by taxonomic groups.

			richness, OTUs	richness, %	abundance, reads	abundance, %
Kinetoplastea			512	100	1570025	100
	Metakinetoplastina		440	85.9	1566578	99.8
		Neobodonida	360	70.3	1544278	98.4
		unknown Neobodonida	172	33.6	322763	20.6
		Neobodo	127	24.8	928002	59.1
		Rhynchomonas	36	7	285766	18.2
		Azumiobodo	13	2.5	7128	0.5
		Rhynchobodo	12	2.3	619	0
		Eubodonida	35	6.8	20864	1.3
		Parabodonida	17	3.3	715	0
		Parabodo	14	2.7	160	0
		Procryptobia	3	0.6	555	0
		Trypanosomatida	28	5.5	721	0
	Prokinetoplastina		72	14.1	3447	0.2
		Perkinsela	41	8	3094	0.2
		unknown Prokinetoplastina	18	3.5	178	0
		Ichthyobodo	13	2.5	175	0



Caption : Figure 1. Rarefaction curves for OTUs: OTU count vs. read number. Slopes calculated for 10 last data points are indicated in the legend on the right. Curves were constructed for the full Kinetoplastea dataset, for the neobodonid clade and for its most abundant sub-groups: Neobodo, Rhynchomonas, and unknown Neobodonida.

64x46mm (300 x 300 DPI)

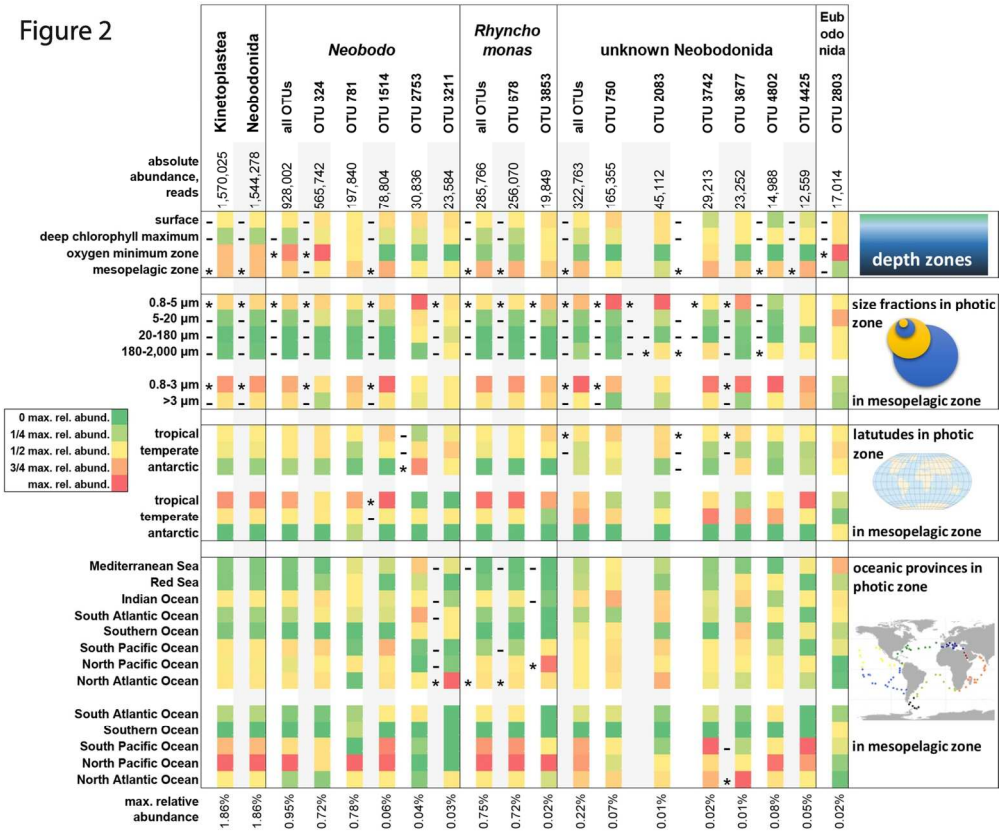


Figure 2. Variation in average kinetoplastid abundance across depth zones, size fractions, and geographical regions. Only most abundant kinetoplastid clades and 14 most abundant OTUs were considered. The bar plots show average relative abundance, with scale at the bottom of each column; and pairs of the minus and asterisk symbols mark significant differences according to one-way ANOVA. Because kinetoplastids were preferentially found in the smallest size fraction of 0.8-5 um and in the mesopelagic zone, geographic variables were considered not only on the whole dataset, but also separately on these subsets. Furthermore, because a different set of size fractions was taken in the mesopelagic zone, the size variability was assessed in this zone separately. The following abbreviations are used: SRF, surface zone; DCM, deep chlorophyll maximum zone; OMZ, oxygen minimum zone; MES, mesopelagic zone; MS, Mediterranean Sea; RS, Red Sea; IO, Indian Ocean; SAO, South Atlantic Ocean; SO, Southern Ocean; SPO, South Pacific Ocean; NPO, North Pacific Ocean; NAO, North Atlantic Ocean.

148x123mm (300 x 300 DPI)

Figure 3

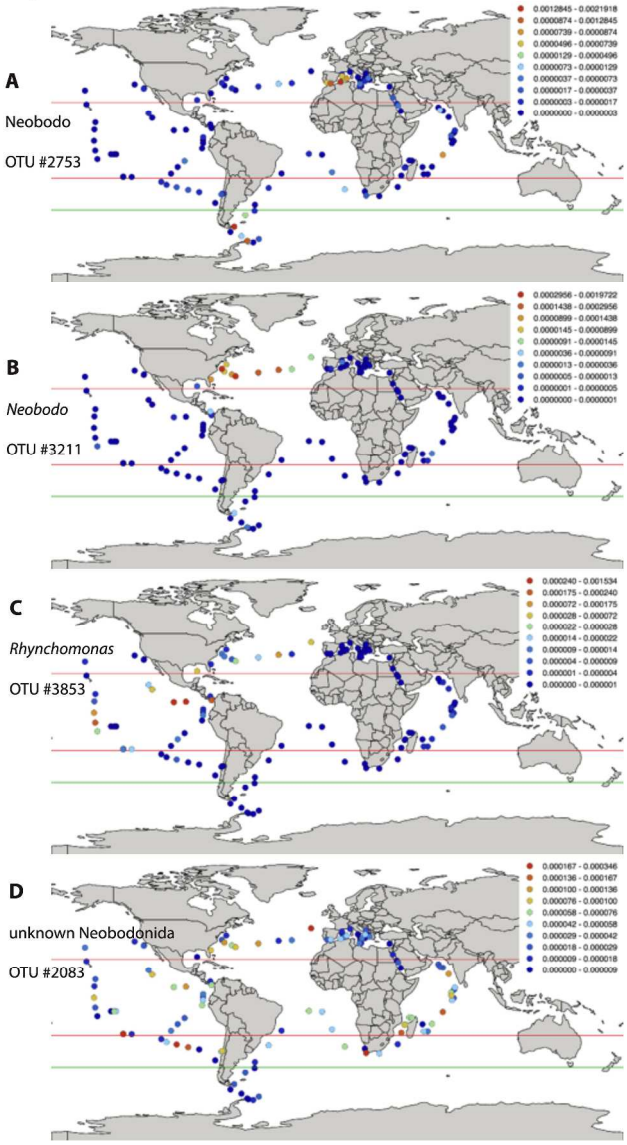


Figure 3. Examples of geographical distribution of abundant OTUs. Two Neobodo OTUs (A, B), one Rhynchomonas OTU (C), and the putatively parasitic OTU #2083 (D) are shown. Full results for these and the other abundant OTUs are presented in Suppl. Figure 1. Relative abundance is color coded, see legends in each panel.

355x681mm (300 x 300 DPI)

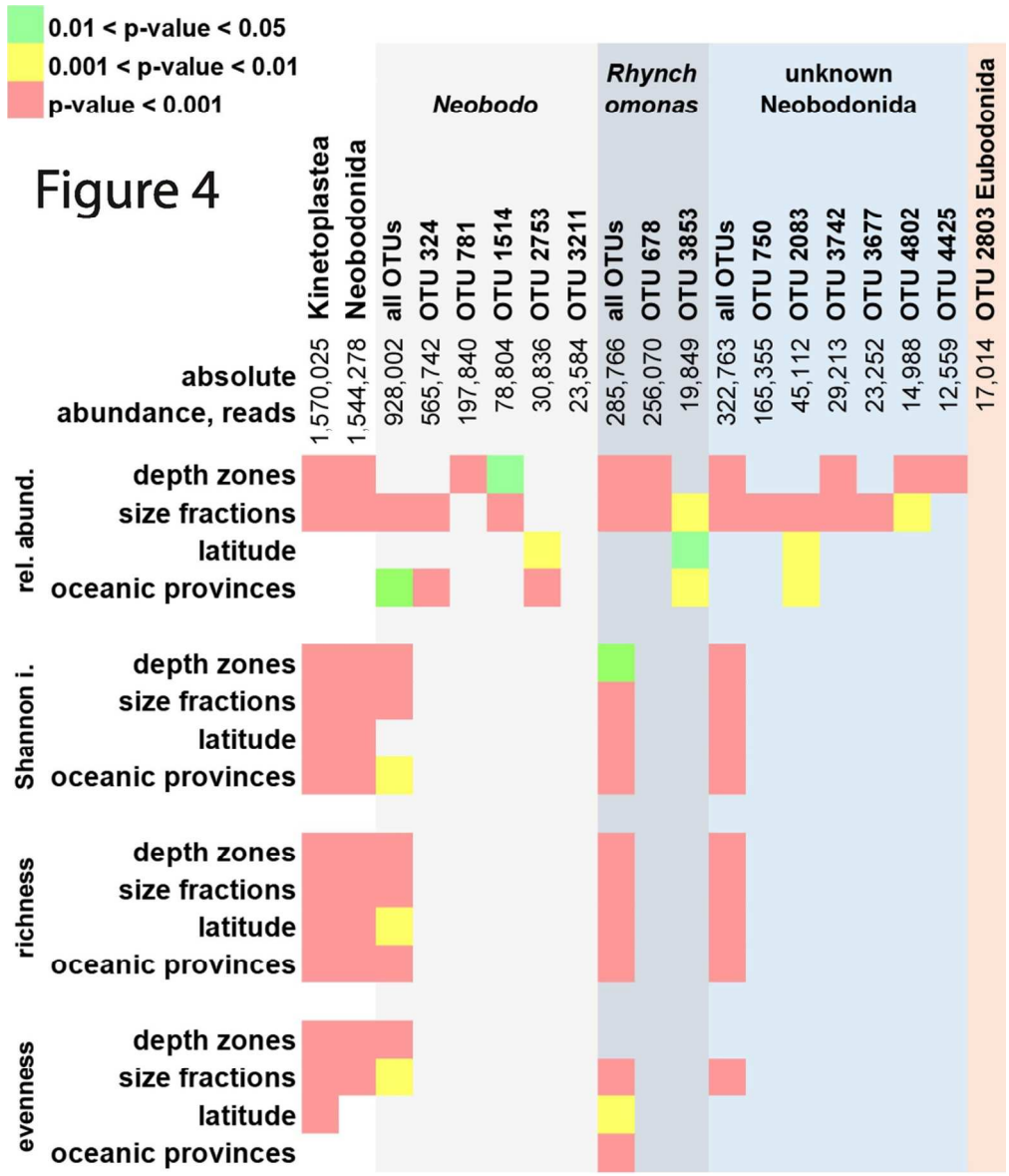


Figure 4. Factors driving abundance and diversity of kinetoplastids. We performed a multi-way ANOVA analysis to determine which variables drive relative abundance and diversity of kinetoplastids and their most abundant sub-clades. The strongest influence we observed was size fractions affecting abundance and diversity. Abundance was also significantly affected by depth, and in case of four OTUs also by oceanic provinces or latitude regions. On the other hand, diversity statistics were significantly affected by all four variables. P-values are coded as follows: full black, p-values < 0.001; chequered, p-values from 0.001 to 0.01; horizontal stripes, p-values from 0.01 to 0.05.

104x122mm (300 x 300 DPI)

Figure 5

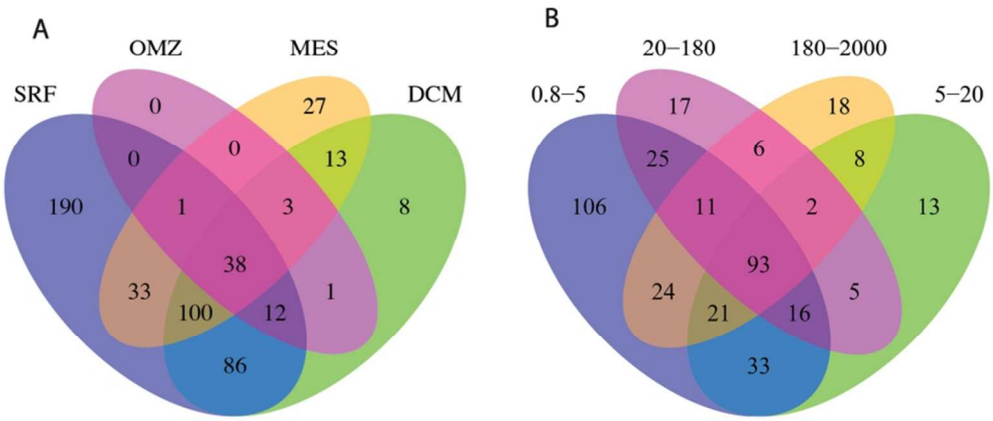


Figure 5. Venn diagrams showing OTU distribution across depth zones (A) and size fractions (B). The following abbreviations are used: SRF, surface zone; DCM, deep chlorophyll maximum; MES, mesopelagic zone; OMZ, oxygen minimum zone.

84x40mm (300 x 300 DPI)

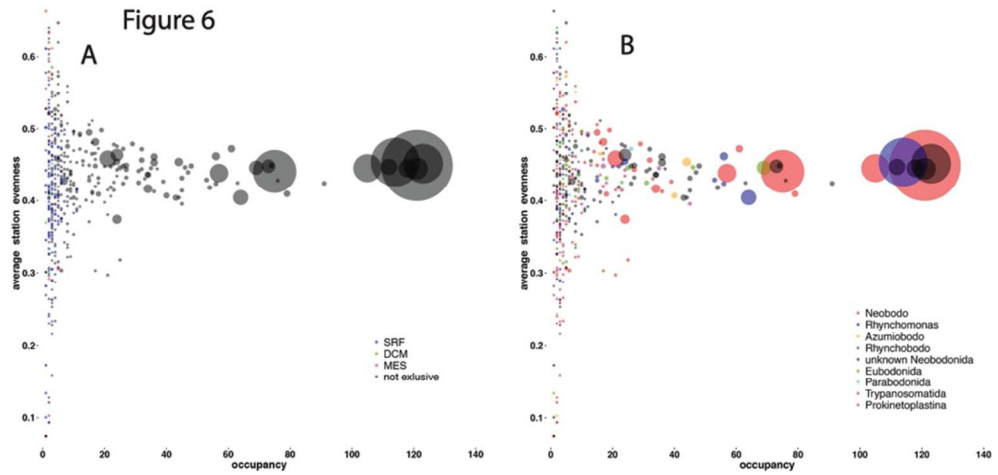


Figure 6. Analysis of cosmopolitan and rare OTUs. Occupancy values, i.e., the number of stations where an OTU was found, are plotted on the x axis, and average station evenness for these stations is plotted on the y axis. Bubble size represents a read count for a given OTU, and OTUs unique to one depth zone (A) or taxonomic group (B) are color-coded according to the legend.

84x40mm (300 x 300 DPI)