environmental microbiology

Environmental Microbiology (2020) **22**(9), 4014–4031 doi:10.1111/1462-2920.15190

Environmental determinants of the distribution of planktonic diplonemids and kinetoplastids in the oceans

Olga Flegontova,[1,2†] Pavel Flegontov,[1,2,6†] Paula Andrea Castañeda Londoño,[1,7] Waldemar Walczowski,[3] Danijela Šantić,[4] Virginia P. Edgcomb,[5] Julius Lukeš[1,8*] and Aleš Horák[1,8*]

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

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

[3]*Institute of Oceanology, Polish Academy of Sciences, Sopot, Poland.*

[4]*Institute of Oceanography and Fisheries, Split, Croatia.*

[5]*Geology and Geophysics Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts.*

[6]*Department of Human Evolutionary Biology, Harvard University, Cambridge, MA.*

[7]*Biocenter, University of Würzburg, Würzburg, Germany.*

[8]*Department of Molecular Biology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.*

Summary

We analysed a widely used barcode, the V9 region of the 18S rRNA gene, to study the effect of environmental conditions on the distribution of two related heterotrophic protistan lineages in marine plankton, kinetoplastids and diplonemids. We relied on a major published dataset (Tara Oceans) where samples from the mesopelagic zone were available from just 32 of 123 locations, and both groups are most abundant in this zone. To close sampling gaps and obtain more information from the deeper ocean, we collected 57 new samples targeting especially the mesopelagic zone. We sampled in three geographic regions: the Arctic, two depth transects in the Adriatic Sea, and the anoxic Cariaco Basin. In agreement with previous studies, both protist groups are most abundant and diverse in the mesopelagic zone. In addition to that, we found that their abundance, richness, and community structure also depend on geography, oxygen concentration, salinity, temperature, and other environmental variables reflecting the abundance of algae and nutrients. Both groups studied here demonstrated similar patterns, although some differences were also observed. Kinetoplastids and diplonemids prefer tropical regions and nutrient-rich conditions and avoid high oxygen concentration, high salinity, and high density of algae.

Received 8 July, 2019; accepted 7 August, 2020. *For correspondence. E-mail ogar@paru.cas.cz, jula@paru.cas.cz; Tel. +420387775409, +420387775403; Fax. +420385310388. †These authors contributed equally.

Introduction

Heterotrophic marine protists have recently received increased attention as an essential component of the plankton, with species number that significantly exceeds that of all photosynthetic eukaryotes (de Vargas *et al*., 2015; Pernice *et al*., 2015; Worden *et al*., 2015). Large protistan predators of the photic zone, such as ciliates, are significant regulators of algal blooms (Calbet and Landry, 2004; Tillmann, 2004). Moving into the deep ocean, protistan bacteriovores, including rhizarians, marine stramenopiles (MAST), ciliates and bodonids (belonging to kinetoplastids), are known to effectively control even dilute prokaryotic communities by removing up to 30% of their biomass (Pachiadaki *et al*., 2016; Rocke *et al*., 2015). Hyper-diverse and abundant parasitoid alveolates belonging to the marine alveolate (MALV) group II can terminate toxic algal blooms (Chambouvet *et al*., 2008) and can control populations of dinoflagellates, rhizarians, metazoans and even other MALVs (Guillou *et al*., 2008; Bråte *et al*., 2012; Massana *et al*., 2014). Heterotrophic protists may thus significantly contribute to pools of dissolved and particulate organic carbon in the ocean, as well as exert controls over overall community structure and dynamics (Šolić and Krstulović, 1994; Šimek *et al*., 1997; Lin *et al*., 2007; Edgcomb, 2016; Šolić *et al*., 2018; Šolić *et al*., 2019). However, we are still far from understanding how the

environment drives their distribution. So far, only a few attempts have been made to explain the global patterns of abundance and diversity of planktonic protists (Lima-Mendez *et al*., 2015; Lima-Mendez *et al*., 2015; Gimmler *et al*., 2016; Zhao *et al*., 2017; Ibarbalz *et al*., 2019).

Diplonemids, the sister clade of kinetoplastids, emerged in a global metabarcoding study of the marine plankton as a prominent group of protists (de Vargas *et al*., 2015). They are divided into four main lineages: Diplonemidae, Hemistasiidae, Eupelagonemidae, and 'deep-sea pelagic diplonemids (DSPD) II'. The latter two groups are 'environmental' and do not include any cultivable species (Lara *et al*., 2009; Okamoto *et al*., 2019; Tashyreva *et al*., 2018). The two former groups are known to scavenge the cytoplasm of plant cells and to cause disease (Larsen and Patterson, 1990), to parasitize crabs, lobsters, and clams (Kent *et al*., 1987; Von Der Heyden *et al*., 2004) and to prey on diatoms, dinoflagellates, haptophytes, and copepods (Elbrächter *et al*., 1996). However, a recent metabarcoding-based study showed that these two cultivable lineages account for just about 2% of planktonic diplonemids, and their genetic signatures are recovered primarily from the photic zone (Flegontova *et al*., 2016). In contrast, the Eupelagonemidae occurs mostly in the mesopelagic zone (Lara *et al*., 2009; Okamoto *et al*., 2019) and represents up to 97% of diplonemids in the plankton (Flegontova *et al*., 2016). So far, they are known only from environmental 18S rRNA sequences and low-coverage single-cell genomes accompanied by light microscopy images (Gawryluk *et al*., 2016; Okamoto *et al*., 2019). Only six out of 100 of the most abundant diplonemid operational taxonomic units (OTUs) were found predominantly in the mesoplankton size fraction (180–2000 μm), suggesting that a symbiotic or parasitic lifestyle with mesoplankton is not widespread among eupelagonemids (Flegontova *et al*., 2016).

Kinetoplastids are one of the most extensively studied protistan lineages. They include trypanosomatids, obligatory endoparasites of mostly terrestrial vertebrates and insects, and a paraphyletic assemblage of basal lineages called 'bodonids' (Lukeš *et al*., 2014). Bodonids occur in marine, freshwater, and soil environments and have diverse lifestyles: a majority are free-living bacteriovores, though some are ecto- and endoparasites or even obligatory intracellular symbionts. Kinetoplastids are subdivided into several clades: the most basal environmental clade composed of sequences from deep-sea hydrothermal vents (López-García *et al*., 2003); a relatively rare Prokinetoplastina clade; and the most diverse Metakinetoplastina, which is further subdivided into Eubodonida, Parabodonida, Neobodonida, and Trypanosomatida (Moreira *et al*., 2004; Von Der Heyden *et al*., 2004). More than 70% of kinetoplastid OTUs and 98% of reads from the Tara Oceans planktonic samples were assigned to the Neobodonida group (Flegontova *et al*., 2018). Marine planktonic kinetoplastids are about 100× less diverse (in terms of OTUs) and abundant relative to diplonemids; nevertheless, they were found in nearly all Tara Oceans samples (Flegontova *et al*., 2018).

In previous global studies of diplonemid and kinetoplastid diversity (Flegontova *et al*., 2016, 2018), deep-sea samples (from depths greater than 200 m) were available from just 32 of 123 locations, mostly in the tropics and sub-tropics. Moreover, only 9 out of 123 sampling stations came from outside of temperate zones, namely from the South Atlantic and Southern Oceans. To close the most significant sampling gaps, we produced metabarcoding data for 57 new samples using the V9 18S rDNA barcode, to be consistent with data from the Tara Oceans survey (see below). This enabled us to investigate the influence of various environmental variables on diplonemid and kinetoplastid abundance, OTU richness, and community structure on global as well as local scales.

Results

Taxonomic composition

We present new metabarcoding data for 57 marine planktonic samples coming from three different oceanic regions. Thirty-seven were collected above the 70th parallel north in the Norwegian and Greenland Seas, and are referred to below as 'Arctic'. Fifteen samples come from two depth transects in the Adriatic Sea, and five samples are from the permanently stratified, anoxic Cariaco Basin in Venezuela. Most samples were collected below the photic zone at depths of down to ca. 1200 m. The basic characteristics of samples are summarized in Table S1. Eukaryotic diversity was investigated using metabarcoding of the V9 region of the 18S rRNA gene. Our protocols for water filtration (except for the choice of size fractions), DNA isolation, V9 barcode amplification and sequencing, OTU clustering, and taxonomic assignment followed the methods of previous studies (de Vargas *et al*., 2015; Flegontova *et al*., 2016, 2018) to enable comparisons and are described in the Methods section.

The taxonomic composition of protist communities in the three geographic regions we examined is summarized in Table 1. The results are consistent with recently published global surveys (de Vargas *et al*., 2015; Pernice *et al*., 2015) that highlight diplonemids as a significant component of the eukaryotic plankton at many sites. The relative richness of diplonemids varies substantially across datasets (from 3.2% in Cariaco Basin to 7.4% in the Arctic) and increases with dataset size (read number). In a combined dataset that also contains reads from

Table 1. Taxonomic profiles of planktonic metabarcoding samples presented here.

	Arctic		Adriatic		Cariaco		Tara		combined dataset		combined dataset
taxonomic group	relative abundance	relative richness	relative abundance	relative richness	relative abundance	relative richness	relative abundance	relative richness	relative abundance	relative richness	rarefaction curve slope
Alveolata	36.71%	19.63%	17.75%	18.63%	33.57%	18.39%	12.15%	14.35%	13.26%	14.48%	5.4E-06
Amoebozoa	0.09%	0.33%	0.00%	0.11%	0.01%	0.23%	0.25%	0.18%	0.24%	0.18%	7.2E-06
Ancyromonadida	0.06%	0.09%	0.00%	0.04%	0.00%	0.08%	0.00%	0.01%	0.00%	0.01%	2.4E-05
Archaeplastida	0.31%	2.21%	0.36%	1.55%	0.12%	1.54%	0.63%	0.79%	0.61%	0.81%	9.3E-06
Cryptista	0.15%	0.51%	0.80%	0.65%	0.17%	0.58%	0.38%	0.25%	0.38%	0.26%	1.0E-05
Diplonemea	0.94%	7.36%	3.63%	5.75%	1.08%	3.17%	2.36%	17.52%	2.32%	17.34%	5.9E-05
Haptista	1.23%	1.61%	1.85%	2.95%	0.40%	1.57%	0.64%	0.43%	0.68%	0.45%	9.1E-06
Kinetoplastea	0.03%	0.39%	0.05%	0.27%	0.01%	0.24%	0.15%	0.14%	0.14%	0.14%	4.7E-06
Obazoa	7.05%	5.04%	29.14%	4.55%	9.05%	4.35%	38.03%	11.41%	36.59%	11.27%	1.1E-06
Picozoa	0.51%	0.35%	0.12%	0.16%	0.03%	0.13%	0.24%	0.14%	0.25%	0.14%	7.5E-06
Rhizaria	11.22%	4.68%	9.61%	4.45%	29.31%	7.37%	16.19%	6.38%	15.96%	6.34%	1.6E-06
Stramenopiles	7.58%	9.87%	3.56%	7.60%	3.05%	8.35%	5.21%	4.01%	5.26%	4.15%	5.6E-06
Telonemia	0.99%	0.48%	1.06%	0.65%	0.31%	0.48%	0.18%	0.10%	0.22%	0.11%	5.8E-06
unknown Eukaryota	31.74%	34.00%	29.20%	34.57%	21.19%	36.75%	13.38%	12.01%	14.35%	12.55%	
other eu- and prokaryotes	1.41%	13.46%	2.86%	18.08%	1.70%	16.77%	10.22%	32.29%	9.74%	31.77%	
Eupelagonemidae	99.69%	96.63%	98.19%	93.77%	99.88%	98.65%	99.61%	98.81%	99.58%	98.81%	
DSPDII	0.30%	2.90%	1.24%	5.86%	0.00%	0.53%	0.37%	1.11%	0.39%	1.10%	
Hemistasia	0.00%	0.07%	0.00%	0.00%	0.00%	0.26%	0.00%	0.00%	0.00%	0.00%	
Diplonema	0.01%	0.34%	0.57%	0.37%	0.11%	0.53%	0.00%	0.02%	0.01%	0.03%	
Flectonema	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	
Rhynchopus	0.00%	0.07%	0.00%	0.00%	0.00%	0.00%	0.02%	0.05%	0.02%	0.05%	
unknown Prokinetoplastina	0.04%	3.85%	0.00%	0.00%	0.20%	3.45%	0.09%	8.65%	0.09%	8.46%	
Ichthyobodo	0.00%	0.00%	0.03%	2.70%	0.00%	0.00%	0.00%	0.38%	0.00%	0.37%	
Perkinsela	0.02%	1.28%	0.00%	0.00%	1.41%	3.45%	0.10%	4.32%	0.10%	4.23%	
unknown Metakinetoplastina	36.95%	61.54%	85.12%	64.86%	93.17%	72.41%	26.43%	46.05%	26.84%	46.32%	
Azumiobodo	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%	1.32%	0.04%	1.29%	
Dimastigella	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.19%	0.00%	0.18%	
Neobodo	23.44%	15.38%	1.42%	10.81%	0.00%	0.00%	54.91%	18.80%	54.40%	18.75%	
Rhynchobodo	0.02%	2.56%	0.00%	0.00%	0.00%	0.00%	0.04%	2.63%	0.04%	2.57%	
Rhynchomonas	39.06%	6.41%	12.95%	16.22%	4.42%	10.34%	16.64%	6.95%	16.77%	6.80%	
Parabodo	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.75%	0.00%	0.74%	
Procryptobia	0.11%	1.28%	0.00%	0.00%	0.00%	0.00%	0.03%	0.56%	0.03%	0.55%	
Eubodonida	0.04%	3.85%	0.00%	0.00%	0.80%	10.34%	1.67%	6.02%	1.65%	5.88%	
Trypanosomatida	0.32%	3.85%	0.48%	5.41%	0.00%	0.00%	0.03%	3.38%	0.04%	3.86%	
	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	

In the upper part, relative abundance and richness of main eukaryotic lineages are shown for each dataset. For the combined dataset, we also show slopes of rarefaction curves reflecting the saturation of observed diversity. Detailed taxonomic breakdown of abundance and richness statistics is also shown for diplonemids (the middle part) and kinetoplastids (the bottom part).

the Tara Oceans survey (de Vargas *et al*., 2015; Flegontova *et al*., 2016), diplonemids are the most diverse (relative richness 17.3%) and the fifth most abundant (relative abundance 2.3%) major eukaryotic phylum (Table 1). Their diversity in the three newly collected datasets remains unsaturated but reaches saturation when combined with the Tara Oceans dataset (Table 1). Diplonemids occupy the fifth place by relative abundance in the combined, Tara Oceans, and Cariaco Basin datasets, the fourth place in the Adriatic (reaching 3.6%), and the seventh place in the Arctic (0.94%, below Telonemia and Haptista). The most prominent diplonemid clade in all samples and datasets is the eupelagonemids. The relative richness and abundance of the remaining lineages are negligible (<<1%). The only exception is an environmental clade DSPD II, which makes up 1.2% of all diplonemid reads and 5.9% of diplonemid OTUs in the Adriatic dataset (Table 1).

Kinetoplastids account for a much smaller share of eukaryotic reads and OTUs as compared to diplonemids in all datasets presented here. Their relative abundance and relative richness reached 0.14% in the combined dataset (Table 1). Three kinetoplastid taxa: *Neobodo* (54% of kinetoplastid reads/19% of kinetoplastid OTUs in the combined dataset), unknown Metakinetoplastina (27/46%) and *Rhynchomonas* (17/7%) dominate our datasets while the rest of the kinetoplastids constitutes a negligible fraction. *Neobodo* is most prominent in the Tara Oceans and Arctic subsets, and almost non-existent in the Adriatic and Cariaco Basin subsets (Table 1). *Rhynchomonas* is present in all samples but is more abundant in the Arctic and Tara Oceans subsets.

To investigate the variability of diplonemid V9 region in the context of the whole 18S rDNA sequence, we amplified and clone- sequenced diplonemid 18S rDNA from eight mesopelagic samples. Eighty-five of 104 diplonemid sequences we got were unique, and polymorphic sites were spread evenly across all variable 18S rDNA domains.

Summary statistics: relative abundance and OTU richness

Similar to other heterotrophic protists, the relative abundance and richness of diplonemids and kinetoplastids remain low in the photic zone, which is typically

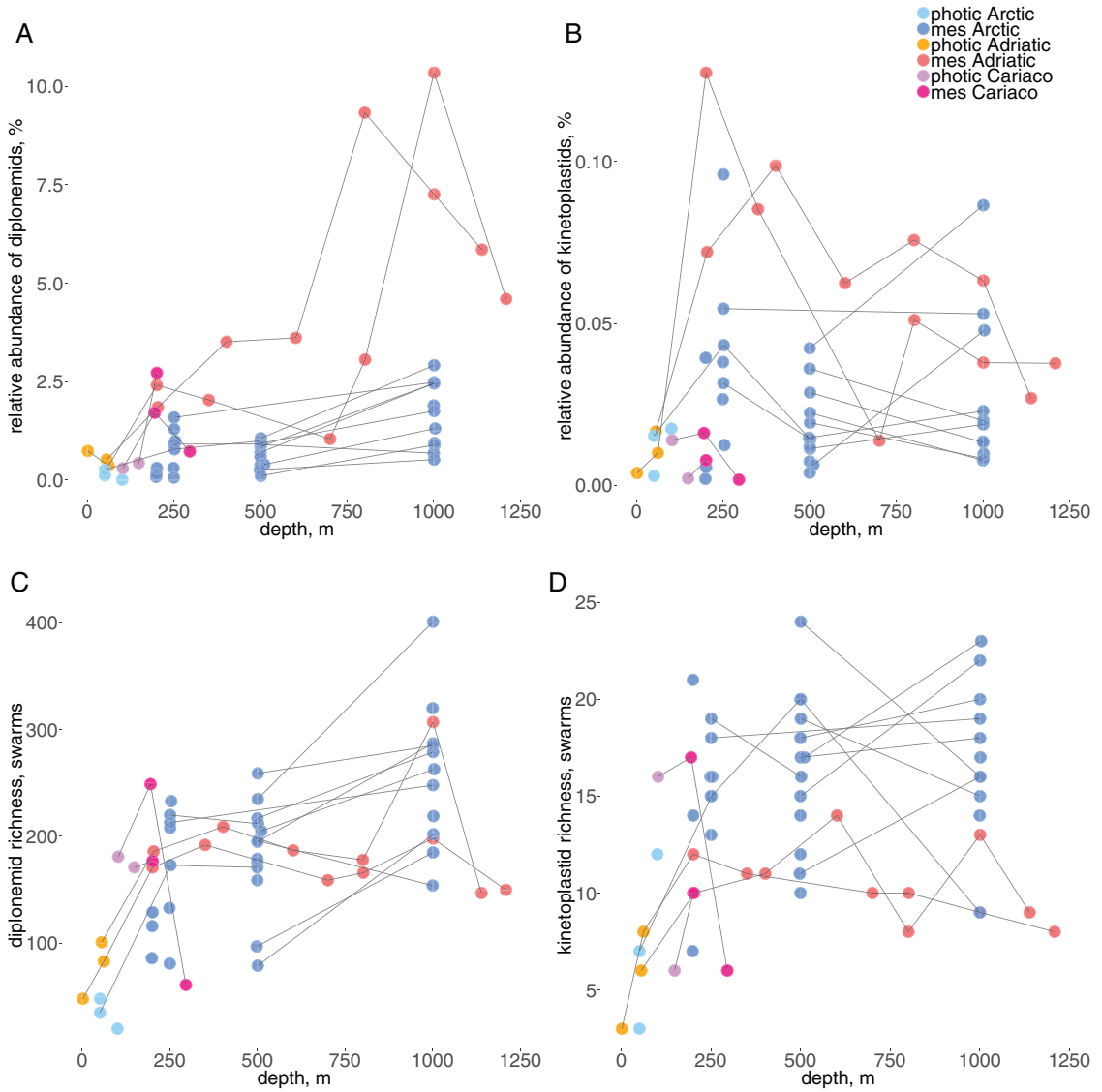

Fig. 1. Distribution of diplonemid (A, C) and kinetoplastid (B, D) relative abundance (A, B) and richness (C, D) across the sampling depths. Each dot represents a single sample colour-coded according to the sampling area (Arctic, Adriatic, and Cariaco) and sampling depth (photic and mesopelagic). Only samples generated in this study are shown. Two or more samples coming from a single sampling site are connected by lines to illustrate trends.

dominated by phototrophs (de Vargas *et al*., 2015). Because of the limited mesopelagic sampling in the Tara Oceans dataset, we aimed to gather additional samples from this zone to obtain a more detailed picture of the vertical distribution of both kinetoplastids and diplonemids. Although only the Adriatic transects have a detailed vertical sampling, it seems that both diplonemid diversity and abundance peak at around 750–1000 m and then drop below these depths (Fig. 1A and C). For kinetoplastids, the differences are less pronounced, and the patterns less obvious. At most of the sampling sites, their abundance peaked between 250–500 m depth (Fig. 1Bs). We analysed just five samples from one location in the Cariaco Basin, sampled during different seasons and at different depths (Supporting Information Table S1). One sample from a depth of 296 m was completely anoxic (Supporting Information Table S1), and it demonstrated the lowest relative abundance of kinetoplastids and the lowest richness of both kinetoplastids and diplonemids among all new mesopelagic samples presented here. However, diplonemid abundance in this sample was not exceptionally low (Fig. 1A).

Next, we wanted to address the effect of environmental variables on the distribution of both groups in the broader context of 57 new samples and 229 Tara Oceans

samples from the photic and mesopelagic zones (http://taraoceans.sb-roscoff.fr/EukDiv/). For this purpose, all size fractions between 0.8 and 2000 μm in the Tara Oceans dataset were merged for each location and depth. For our newly generated datasets, only the following quantitative variables were available: latitude, longitude, depth, water density (σ_θ), temperature, salinity, and oxygen concentration. In contrast, for the Tara Oceans dataset, 47 quantitative variables were known, albeit some with high missing data rates (Supporting Information Table S1).

In the photic zone, diplonemid relative abundance and richness respond in a similar way to most environmental variables. On the other hand, in the mesopelagic zone, diplonemid relative abundance and richness respond in different ways to many variables, such as depth, latitude, and temperature (results not shown). The observed discrepancy may be caused by the very low total concentration of eukaryotic cells in mesopelagic samples, and perhaps diplonemids happen to be more abundant there than other eukaryotes. However, it may be misleading to say that diplonemids prefer these conditions. Moreover, since relative barcode abundance can be affected by well-known amplification biases across clades (Giner *et al*., 2016, Kelly *et al*., 2019) and by the variable copy number of rRNA genes in different organisms, it may be a poor proxy for cell counts. Thus, we decided to pay more attention to the richness and not to the relative abundance statistics.

The richness and relative abundance of diplonemids and kinetoplastids were lower in the newly generated datasets as compared to the Tara Oceans samples but on a par with Tara Oceans photic samples from the Antarctic (Supporting Information Fig. S1). It is not surprising that our Arctic samples and Tara Oceans Antarctic samples demonstrated similar levels of diplonemid/kinetoplastid relative abundance and diversity. Also, although there are 19 samples present with the oxygen concentration below 4 mg/l in the original Tara Oceans dataset, the conditions in the Cariaco Basin may be unique enough to distinguish it from the rest of the samples. However, it is rather unexpected that the Adriatic samples, especially those from the mesopelagic zone, differ substantially from the Tara Oceans samples from the same region and similar environments. Since this result may be attributed to ecosystem differences or biases arising from differences in sampling protocols, we first explored how the biological variables depend on the environmental variables in the full dataset and then on the Tara Oceans samples only.

Linear regression is not an appropriate method for ecological studies where non-monotonic response curves are often observed. Therefore, we approximated the distributions using generalized additive models (GAM). Results for diplonemid richness depending on six environmental variables measured across all datasets are shown in Fig. 2. The models fitted to the distributions for latitude, oxygen concentration, and temperature remain

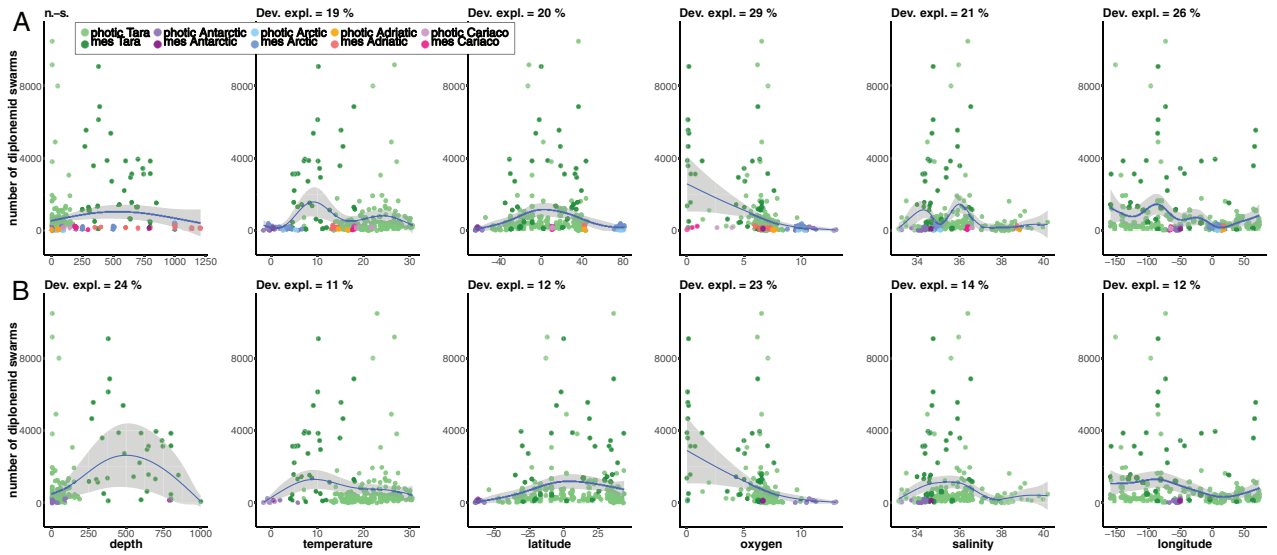

Fig. 2. Scatterplots showing the distribution of diplonemid richness across six environmental variables. Samples are represented by dots coloured according to the dataset and depth zone. Blue lines represent the GAM smooth trends and grey ribbons, the corresponding 95% confidence intervals of diplonemid richness predicted by the GAMs. If the *P*-value of smooth terms is <0.05, the deviance explained is shown in the upper right corner of the plot. Otherwise, it is written 'n.-s.' (nonsignificant) in the upper right corner of the plot.
A. Scatterplots showing the distribution of diplonemid richness for the combined dataset.
B. Scatterplots showing the distribution of diplonemid richness for the Tara Oceans dataset only.

nearly the same when our new samples are dropped, whereas the models fitted to the distributions for depth, longitude, and salinity change only slightly. These results are reassuring since they imply that the sampling bias is not severe.

Since temperature weakly correlated with depth (Supporting Information Fig. S2C), and the mesopelagic samples were colder than the photic ones, we observe two peaks of diplonemid richness versus temperature (Fig. 2A). Latitude and oxygen concentration are well correlated (Supporting Information Fig. S2C), and we observe a substantial decrease in diplonemid richness at high latitudes and high oxygen concentrations. It is of particular interest that diplonemid richness remains very high in nearly anoxic samples (oxygen concentration from 0 to 1 mg/l). Analysis of the original Tara Oceans dataset implies the existence of a salinity optimum for diplonemids between ~34 and ~36 (Fig. 2B). However, the inclusion of numerous new Arctic samples, all having salinity within a narrow range of 34.7–35.1, breaks this salinity optimum into two peaks since all Arctic and Antarctic samples demonstrated low diplonemid richness. Samples from the Mediterranean and Red Seas, known to have high salinity, are poor in diplonemids. The longitudinal distribution probably reflects the effect of salinity since numerous Tara Oceans samples from the Mediterranean and Red Seas fall into the 'valley' between 0° and 50° east (Fig. 2B). Kinetoplastid richness shows a pattern very similar to that of diplonemids: there is a more substantial effect of temperature on kinetoplastids in the photic zone as compared to diplonemids in the same zone; a stronger effect of latitude, and smaller effects of salinity and oxygen concentration on kinetoplastids as compared to diplonemids (Supporting Information Fig. S2).

These results suggest that even this limited set of analysed environmental variables influence diplonemid and kinetoplastid richness and distribution; yet, we cannot exclude the possibility that there are also other important environmental parameters with a strong impact on those protistan groups. Therefore, we have limited further analyses of impacts of environmental factors on the distribution of these taxa to only the Tara Oceans dataset, for which 47 variables are available. We fitted GAMs to the distributions (diplonemid richness vs. an environmental variable) with no outliers on the environmental variable axis removed, or with extreme outliers removed (see Methods section). Next, variables have been selected that satisfied the following criteria: (i) a statistically significant GAM fit for both distributions, with and without outliers (P-value < 0.05); and (ii) deviance explained by the GAM exceeds 5% on the distribution without outliers. As a result we got 23 environmental variables that each explained between 6% and 26% of variance in diplonemid richness, and all six variables discussed above were among them. The scatterplots for these 23 environmental variables and all four biological variables are shown in Supporting Information Fig. S3, and a correlation dendrogram for all 47 variables is shown in Supporting Information Fig. S4A.

The best-fitting model was generated for diplonemid richness versus chlorophyll A concentration: a negative dependence was observed, and 26% of the variance was explained (Supporting Information Fig. S3), with a similar trend observed for the other three biological variables (diplonemid abundance, kinetoplastid abundance and richness). This result suggests that most diplonemid and kinetoplastid OTUs are not associated with the same habitat (photic zone) where algae are found. An alternative explanation is that high chlorophyll concentrations are caused by blooms when one or few algal species predominate, and we do not expect many diplonemid OTUs to be associated with those few algae. Chlorophyll A concentration correlates with backscattering at 470 nm, with partial optical density at 660 nm (Fig. S4A), and all these variables show well-fitting GAMs for diplonemid richness (Fig. S3) and very similar curves.

Fluorescence is correlated with nitrate and nitrite concentration, and both variables may influence diplonemid richness, but not kinetoplastid richness or abundance. The concentration of silicate does not correlate with other environmental variables, and its optimum for diplonemid and kinetoplastid richness seems to be reached in the mesopelagic samples. On the other hand, we see a nitrate concentration interval (measured at the 5 m depth) that is optimal for both diplonemid and kinetoplastid richness throughout the water column (i.e., even in the water mass below in the mesopelagic zone). This might reflect a possibility that specific nitrate concentrations at the surface support the algal density, which drives the food chain throughout the whole water column through sinking organic material. We also observe weakly fitting curves for calcite saturation, total carbon concentration, and carbonate concentration. The latter two variables are strongly correlated with each other, and the former is weakly correlated with them. All these variables are negatively correlated with diplonemid richness and kinetoplastid richness, for reasons that remain unclear. Remarkably, all the mesopelagic samples show zero values of those three variables, but the effects are also visible in the photic zone. Seafloor depth possibly influences diplonemid richness in the water column with a unimodal curve.

Community structure

To move beyond the four summary statistics discussed above and to study the diplonemid and kinetoplastid community composition in the context of geography and

environmental variables, we applied an ordination method, the redundancy analysis (RDA), which is a linear dimensionality reduction method that takes environmental variables into account. Although Canonical Correspondence Analysis is more appropriate for unimodal distributions, it performs poorly on data with a very large number of dimensions. We took advantage of the combined dataset, including 57 samples sequenced in this study and 230 Tara Oceans samples.

Besides, we performed RD analyses also for various subsets: photic, mesopelagic, Arctic, Mediterranean, and low-oxygen (<5 mg/l). An OTU abundance cut-off of 10 reads was chosen for diplonemids to remove a large group of extremely rare OTUs, and no abundance cut-off was applied for kinetoplastids. Thus, the RD analyses for diplonemids were based on Hellinger-transformed read counts for 21 147 most abundant OTUs among 66 870 diplonemid OTUs. For kinetoplastids, the analysis was based on all known 539 OTUs. The RDA results are summarized below for each sample subset, starting from the largest one.

Combined sample set. The largest sample set includes 287 samples, and the analysis takes into consideration 24 quantitative environmental variables (see Methods section). An RDA plot is shown in Fig. 3 (constrained variance 42%, adjusted r^2 = 0.36). In all datasets (Tara Oceans, Arctic, Adriatic, Cariaco basin), the mesopelagic samples are separated from the photic zone samples. The most important variables, which distinguish the photic and mesopelagic diplonemid communities, are depth, temperature, latitude, absolute latitude, and fluorescence. Samples within both photic and mesopelagic clusters are scattered along a gradient of correlated environmental variables (chlorophyll A concentration, optical density at 660 nm, backscattering at 470 nm, Supporting Information Fig. S4A), and along gradients of daylight hours and silicate concentration. We believe this chlorophyll A gradient reflects the concentration of algae. The Antarctic photic samples occupy a terminal position within the photic sample cline. Salinity, longitude, oxygen concentration, and other variables are less important.

Distribution of kinetoplastid communities shows a less clear pattern (18 quantitative environmental variables selected, constrained variance 30%, adjusted r^2 = 0.25), and separation of various datasets is more prominent than depth stratification (Supporting Information Fig. S5). Clusters of the Tara Oceans mesopelagic and Tara Antarctic photic samples are less pronounced and largely overlap with the Tara Oceans photic cluster. The cline of Tara Oceans samples in the RDA space is parallel to gradients of nitrite and nitrate concentration at 5 m, and

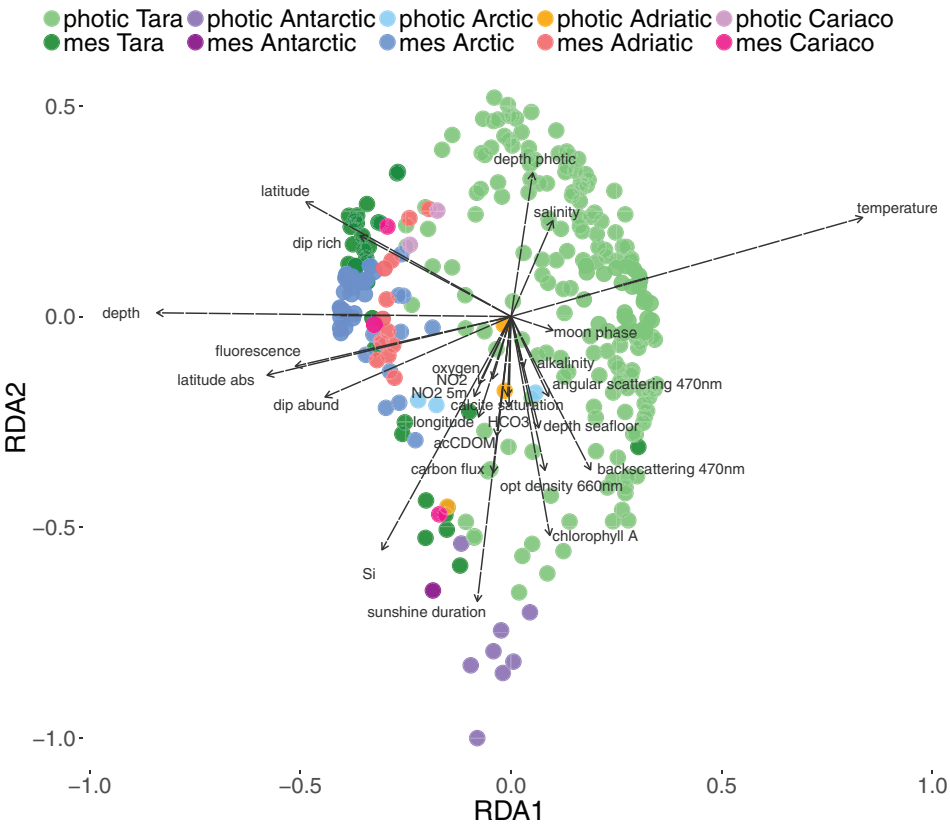

Fig. 3. Impact of environmental variables on the diplonemid community structure in the combined dataset. The RDA plot computed on 21 147 diplonemid OTUs from 287 samples, using 24 selected environmental predictor variables and two biological predictor variables (richness and relative abundance of diplonemids). The adjusted r^2 of the analysis is 0.359 (0.417 unadjusted). Samples are represented by dots coloured according to the dataset and depth zone. Arrows correspond to the predictor variables.

also parallel to gradients of kinetoplastid richness and relative abundance. In other words, these kinetoplastid communities are (either directly or indirectly) sensitive to nitrite/nitrate concentration. Our additional 57 samples are distributed along the gradients of salinity and a pair of very well correlated variables: backscattering at 470 nm and partial backscattering. According to this RDA, oxygen concentration has possibly more influence on the kinetoplastid communities than on diplonemids. In summary, patterns of community structure look very similar for diplonemids and kinetoplastids, and five variables that were measured for almost all samples (depth, latitude, temperature, oxygen concentration, and salinity) emerged among the most influential variables shaping the community structure of both euglenozoan groups.

Photic zone samples. This data set includes 201 photic zone samples, and the analysis considers 22 selected quantitative environmental variables (constrained variance 39%, adjusted r^2 = 0.31) (Fig. 4A). Photic zone diplonemid communities are clustered according to depth, and in the photic subset, depth is correlated only with photosynthetic active radiation (PAR) (Supporting Information Fig. S4B). Thus, the surface samples are differentiated from the deep chlorophyll maximum samples. Another gradient perpendicular to depth is formed by a group of variables (latitude, temperature, oxygen concentration, sunshine duration, nitrite and nitrate concentration, chlorophyll A concentration). These variables thus shape the diplonemid photic communities independently of depth. The Antarctic samples are again well differentiated from the other photic samples.

Kinetoplastid communities in photic zone samples (19 quantitative environmental variables selected, constrained variance 27%, adjusted r^2 = 0.19) do not appear to be influenced by the measured environmental variables to the same extent as their diplonemid counterparts. There is no evident structuring of the kinetoplastid community with depth. The 'latitudinal' gradient (absolute latitude), oxygen concentration, daylight hours, ammonium concentration at 5 m, temperature, and nitrite/nitrate concentration appear to influence kinetoplastid distribution, but to a lesser extent than in diplonemids (Supporting Information Fig. S6A).

Mesopelagic zone samples. We observed a very strong batch effect in the combined mesopelagic dataset for diplonemids. There were three non-overlapping clusters: (i) Tara Oceans cluster, (ii) Arctic cluster, and (iii) Adriatic/Cariaco cluster (results not shown). We cannot exclude a possibility that the clustering is an artefact of difference in size fractions studied, in sampling, DNA extraction, and amplification protocols. Therefore, we only looked at RDA for the Tara Oceans mesopelagic

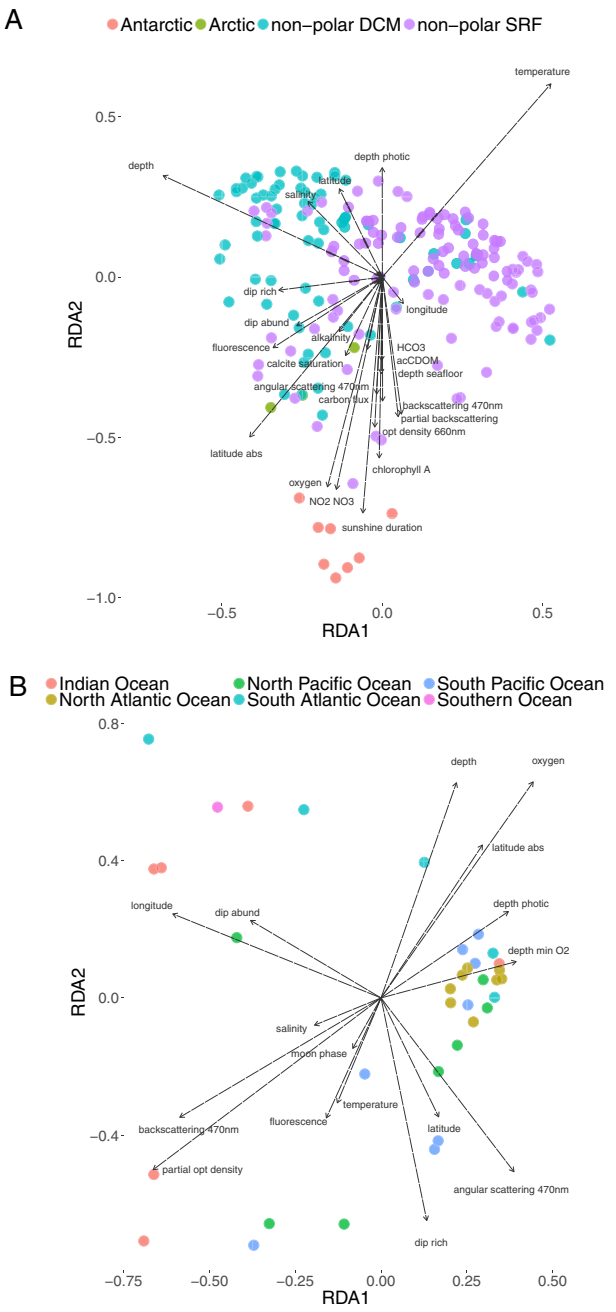

Fig. 4. Impact of environmental variables on the community structure of diplonemids from photic (A) and mesopelagic (B) depth zones.
A. The redundancy analysis (RDA) plot computed on 16 281 diplonemid OTUs from 201 photic samples, using 22 selected environmental predictor variables and two biological predictor variables (richness and relative abundance of diplonemids). The adjusted r^2 of the analysis is 0.312 (0.394 unadjusted). Samples are represented by dots coloured according to geographical regions and depth zones for non-polar samples. Arrows correspond to the predictor variables.
B. An RDA plot computed on 16 338 diplonemid OTUs from 35 Tara Oceans mesopelagic samples, using 11 selected environmental predictor variables and two biological predictor variables (richness and relative abundance of diplonemids). The adjusted r^2 of the analysis is 0.372 (0.668 unadjusted). Samples are represented by dots coloured according to biogeographical provinces. Arrows correspond to the predictor variables.

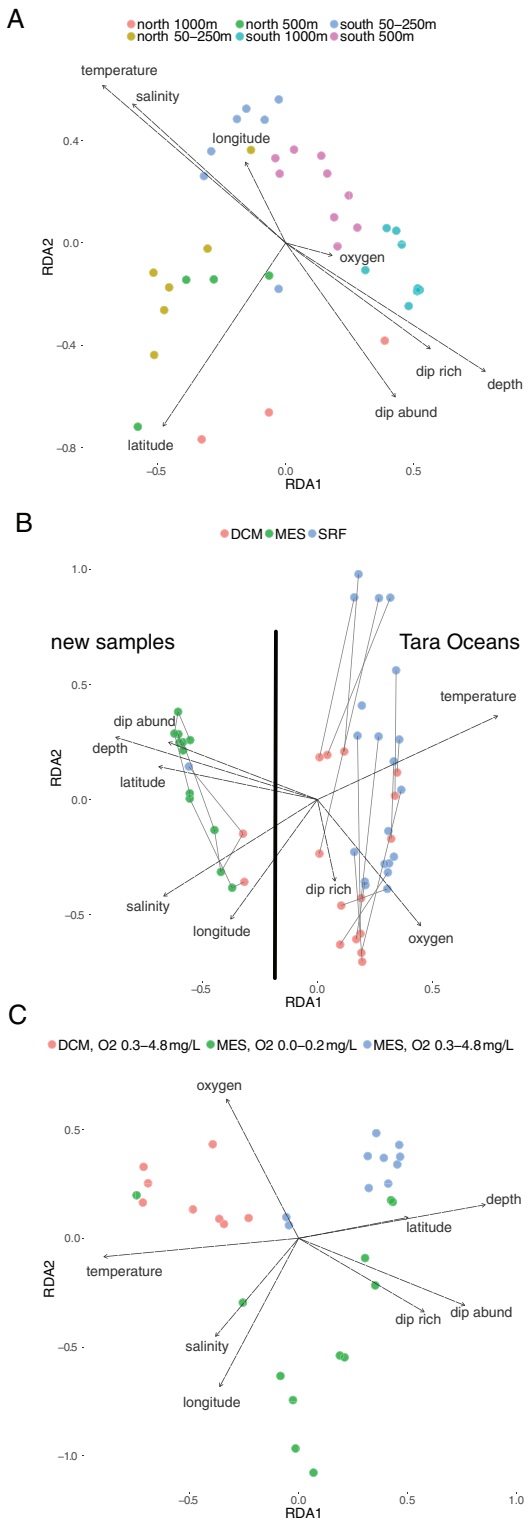

Fig. 5. Impact of environmental variables on the community structure of diplonemids from the Arctic dataset (A), from Mediterranean samples (B), and samples with oxygen concentration below 5 mg/l (C).
A. An RDA plot computed on 700 diplonemid OTUs from 37 Arctic samples, using six environmental and two biological (richness and relative abundance of diplonemids) predictor variables. The adjusted r^2 of the analysis is 0.456 (0.577 unadjusted). Samples are represented by dots coloured according to latitude and depth. Arrows correspond to the predictor variables.
B. An RDA plot computed on 2 552 diplonemid OTUs from 15 new Adriatic samples and 34 Tara Oceans Mediterranean samples (separated by the bold line), using six environmental and two biological (richness and relative abundance of diplonemids) predictor variables. The adjusted r^2 of the analysis is 0.382 (0.485 unadjusted). Samples are represented by dots coloured according to depth zones. Two or more samples coming from a single sampling site are connected by a line. Arrows correspond to the predictor variables.
C. An RDA plot computed on 12 793 diplonemid OTUs from 30 samples with oxygen concentration below 5 mg/l, using six environmental and two biological (richness and relative abundance of diplonemids) predictor variables. The adjusted r^2 of the analysis is 0.325 (0.511 unadjusted). Samples are represented by dots coloured according to depth zones and oxygen concentration for mesopelagic samples. Arrows correspond to the predictor variables.

samples (35 samples, 14 variables selected, constrained variance 67%, adjusted r^2 = 0.37) (Fig. 4B). We observed clustering according to the biogeographical provinces, and strong effects of depth, latitude, and oxygen concentration, with the latter two being correlated variables (Supporting Information. Fig. S4C). Temperature and salinity have only a weak effect.

Similar clustering according to the datasets was also pronounced in the case of kinetoplastid mesopelagic communities (results not shown). If only the Tara Oceans mesopelagic samples are considered (11 variables selected, constrained variance 52%, adjusted r^2 = 0.22), we observe two separate clusters (Indian and North Atlantic Oceans, which form a longitudinal gradient), while the other biogeographical provinces are intermixed (Supporting Information Fig. S6B). Effects of the other environmental variables are not substantial.

Arctic samples. The Arctic diplonemid communities (37 samples, six variables, constrained variance 58%, adjusted r^2 = 0.46) are shaped by two orthogonal gradients (Fig. 5A). The first comprises depth, temperature, and salinity. All three variables are correlated in the Arctic dataset (Supporting Information Fig. S4D). The second gradient includes latitude. For the Arctic kinetoplastid communities (six variables, constrained variance 46%, adjusted r^2 = 0.31), the latitudinal gradient is also prominent, but the depth/temperature/salinity gradient is diffuse, e.g., there are no sample clusters well-separated by depth (Supporting Information Fig. S7A).

Mediterranean samples. In the case of Mediterranean diplonemid communities (49 samples, six variables, constrained variance 49%, adjusted r^2 = 0.38), we observe a strong stratification by depth and a batch effect (all new samples are differentiated from the Tara Oceans samples) (Fig. 5B). In the case of Mediterranean kinetoplastid communities (constrained variance 42%, adjusted r^2 = 0.30), the batch effect is also apparent. Instead of

depth stratification, longitudinal stratification is prominent (Supporting Information Fig. S7B).

Low-oxygen samples. Diplonemid communities in low-oxygen samples (30 samples, six variables, constrained variance 51%, adjusted $r^2 = 0.33$) are stratified by depth and oxygen concentration (Fig. 5C). Kinetoplastid communities in these samples (six variables, constrained variance 49%, adjusted $r^2 = 0.29$) are not well stratified according to the variables explored, except for longitude (Supporting Information Fig. S7C).

Discussion

Until recently, the scientific community focused mostly on parasitic kinetoplastids due to their impact on human health (El-Sayed *et al*., 2005) or on photosynthetic euglenids as a promising biotechnology target (Rodríguez-Zavala *et al*., 2010; O'Neill *et al*., 2015; Ebenezer *et al*., 2019) The rest of the euglenozoans did not receive much attention. However, recent reports of high local (Lara *et al*., 2009; Sauvadet *et al*., 2010; Scheckenbach *et al*., 2010; Edgcomb *et al*., 2011; Salani *et al*., 2012; Mukherjee *et al*., 2015) and global (de Vargas *et al*., 2015; Flegontova *et al*., 2016, 2018) diversity and abundance of free-living kinetoplastids and diplonemids brought these so-far overlooked euglenozoans into the spotlight. Most of these recent results show that diplonemids and free-living kinetoplastids are present throughout the water column. Unfortunately, the largest recent global metabarcoding datasets either used a methodology that underestimated diplonemid and kinetoplastid counts (Pernice *et al*., 2015) or contained a limited sampling of the mesopelagic zone, polar regions, and oxygen-deficient habitats (de Vargas *et al*., 2015; Flegontova *et al*., 2016).

Our main goals were to close these sampling gaps and to estimate the impact of the environment on the diplonemid and kinetoplastid communities. The first step was to analyse biological diversity and relative abundance of both groups. For this aim, we have analysed the taxonomic composition of the studied samples and showed it to be consistent with the general picture revealed by global metabarcoding studies (Massana *et al*., 2014; Pernice *et al*., 2015; de Vargas *et al*., 2015). In addition to other major planktonic lineages, such as metazoans, dinoflagellates, rhizarians, and stramenopiles, diplonemids were consistently recovered among the top taxonomic groups in all new sample datasets provided by this study (Table 1). Diplonemid diversity remains unsaturated in the newly collected sample sets and the broad set of 334 Tara Oceans samples from the photic zone (de Vargas *et al*., 2015). It becomes saturated in an expanded set of 850 Tara Oceans samples from both the photic and mesopelagic zones (Flegontova *et al*., 2016) and when both our new and Tara Oceans samples are combined (Table 1). These results could be explained by the nature of the diplonemid OTU abundance curve showing an extremely long 'tail' of rare OTUs. Flegontova *et al*. (2016) showed that about 93% of diplonemid reads could be assigned to only about 0.2% of OTUs. This predominance of just a few abundant OTUs is not a unique feature of diplonemids. Similar patterns are observed (to various extents) in diatoms, pelagophytes, dinoflagellates, and some other protists (Stoeck *et al*., 2010; Logares *et al*., 2014; Keeling and del Campo, 2017).

Similar results, confirming that the 18S rDNA of excavates is super-diverse, were obtained during the screening of planktonic biodiversity in Scandinavian lakes (Khomich *et al*., 2017). Excavata, alongside with Ichthyosporea and Bicosoecida, despite lower than average read abundance, demonstrated the lowest ratio of total reads to OTU counts (Khomich *et al*., 2017). We confirmed this result using limited sequencing of full-length 18S rDNA from mesopelagic diplonemids. Thus, diplonemid 18S rDNA hyper-diversity is real, and most likely, it is not attributed to intracellular variability (Gertraud Burger, personal communication).

Our results confirmed the dominant role of eupelagonemids (Okamoto *et al*., 2019) among the planktonic diplonemids (Lara *et al*., 2009; Flegontova *et al*., 2016). The global Tara Oceans dataset revealed that each of the remaining diplonemid lineages (Diplonemidae, Hemistasiidae, and 'DSPDII') contributed approximately 1% of total diplonemid abundance and richness (Flegontova *et al*., 2016). In our 57 new samples, these lineages were negligible (<<1%) except for 'DSPDII' which makes up 1.2% of diplonemid reads and 5.9% of OTUs in the Adriatic dataset. Given the low sample size of the Adriatic dataset, this increased abundance of an otherwise rare diplonemid lineage could be a result of stochasticity in sampling. On the other hand, it could also reflect the presence of local microniches suitable for the DSPDII diplonemids.

The parasitic kinetoplastids belong to the most studied (del Campo *et al*., 2014) and possibly one of the most diverse clades of terrestrial protists (Hamilton *et al*., 2007; Maslov *et al*., 2013; Lukeš *et al*., 2014). On the other hand, kinetoplastids seem to constitute a minor component of the plankton. However, several studies revealed an unexpectedly frequent appearance of these flagellates in FISH-based analyses of the mesopelagic and deeper layers (Morgan-Smith *et al*., 2011), as well as in the hypolimnion of deep freshwater lakes (Mukherjee *et al*., 2015). In the combined dataset presented here (i.e., Tara Oceans and new samples reported in this study), the relative abundance of kinetoplastids is 0.14%.

This number is consistent with several previous local reports from the 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). Same as in these studies, *Neobodo* and *Rhynchomonas* are the most abundant planktonic kinetoplastids in our study. These two genera traditionally belonged to Neobodonida, yet the monophyly of this group is not conclusively supported (Callahan *et al*., 2002; Simpson *et al*., 2002; Simpson and Roger, 2004; Cavalier-Smith, 2016; Yazaki *et al*., 2017). Since we cannot rule out the possibility that neobodonids represent an assemblage of basal paraphyletic lineages within Metakinetoplastina, we decided to omit the term Neobodonida from taxonomic annotations in the current study. Therefore, we classified several abundant 'neobodonid' OTUs as an unknown affiliation within Metakinetoplastina. Together with the genera *Neobodo* and *Rhynchomonas*, they make up the vast majority of the kinetoplastid reads in all datasets. Interestingly, although the unknown Metakinetoplastina group is not prominent in the global Tara Oceans dataset, it dominates kinetoplastid communities in both the Adriatic and Cariaco datasets.

The following environmental variables were available for a majority of the samples: latitude, longitude, depth, water density (σ_θ), temperature, salinity, and oxygen concentration. Out of these, we found that geography and all five environmental variables have a pronounced effect on relative abundance and diversity of diplonemids and kinetoplastids. Geography and all these variables also apparently influence the diplonemid and kinetoplastid community composition. Along with nutrients, most of these variables are generally recognized as essential drivers of planktonic diversity and distribution (Righetti *et al*., 2019; Ibarbalz *et al*., 2019).

Apart from the above-described variables available for all the datasets, among a wider set of 47 variables available for the Tara Oceans dataset, we found some interesting relationships. The most conspicuous was the negative correlation between all four biological variables and chlorophyll A concentration. In contrast, the correlation with overall fluorescence (correlated with nitrite, nitrate, and phosphate concentrations) is positive for diplonemids. Chlorophyll A concentration can be used as a proxy for algal abundance and activity. Our results suggest that neither diplonemids nor kinetoplastids prefer conditions of algal blooms, where the concentration of chlorophyll A is the highest. Therefore, the majority of diplonemids and planktonic kinetoplastids does not seem to be predators or symbionts of the phytoplankton. On the other hand, the environment with high overall fluorescence, to which also other organisms contribute, may represent a post-bloom nutrient-rich environment of high planktonic diversity. We assume that when this sinking organic matter from algal blooms reaches the deep zones of the ocean, the energy released into the environment in the form of dead algal biomass represents a major benefit for diplonemids and kinetoplastids.

Generally, the absolute abundance of heterotrophic protists is the highest in epi- and mesopelagic zones and then gradually drops with increasing depth (Edgcomb, 2016). This trend is interrupted by hotspots of prokaryotic and eukaryotic life along the oxy- and redoxclines (Fenchel and Finlay, 2008; Stock *et al*., 2012). Previous studies have shown that the relative abundance and diversity of both diplonemids and kinetoplastids are much higher in the mesopelagic zone as compared to the photic zone (Flegontova *et al*., 2016, 2018). However, those conclusions were based on only a limited number of mesopelagic samples coming from a single depth per sampling site. Moreover, both kinetoplastids and diplonemids were reported in the deeper bathypelagic and abyssopelagic zones (Countway *et al*., 2007; Lara *et al*., 2009; Sauvadet *et al*., 2010; Scheckenbach *et al*., 2010; Eloe *et al*., 2011; Salani *et al*., 2012). Our analysis of a limited set of sampling sites with more than one mesopelagic depth sampled shows that the peak of kinetoplastid abundance occurs in the upper part of this zone. In contrast, for diplonemids, the peak seems to be shifted to the lower mesopelagic zone and the boundary between the meso- and bathypelagic zone at about 1000 m (Fig. 1). At the community level, we observe differences between the meso- and epipelagic diplonemid (but not kinetoplastid) communities (Fig. 4; Fig. S5), which is in agreement with previous results (Flegontova *et al*., 2016).

Neobodo and *Rhynchomonas*, which represent a vast majority of kinetoplastids in our datasets, are thought to be mostly bacteriovores (Lukeš *et al*., 2014), so their distribution likely reflects the distribution of their prey. Biological information is only available for a handful of diplonemid species (Porter, 1973; Kent *et al*., 1987; Larsen and Patterson, 1990; Elbrächter *et al*., 1996; Von Der Heyden *et al*., 2004; Yabuki and Tame, 2015) out of tens of thousands of OTUs. Moreover, eupelagonemids, which are by far the most abundant, diverse, and therefore the ecologically essential group of diplonemids, remain uncultured with information available only from single-cell genomics (Gawryluk *et al*., 2016; Okamoto *et al*., 2019). While some of their representatives seem to be associated with large plankton and may thus represent symbionts or parasites (Flegontova *et al*., 2016), we see no clear pattern in the global ocean interactome (Lima-Mendez *et al*., 2015) that could provide any hint. All our attempts to explain the vertical distribution of

diplonemids would, at this stage, be pure speculation, which we will refrain from.

The effects of the environment are probably best studied for oxygen concentration. Steep oxygen gradients are common in aquatic environments, including the benthos. Oxygen-depleted and anoxic water columns occur either permanently or seasonally in lakes, productive coastal seas with a limited circulation of water, such as fjords (Fenchel and Finlay, 2008), and in large open-ocean oxygen-depleted zones. Gradients of oxygen concentration allow for rapid changes in prokaryotic and eukaryotic communities and often result in hotspots of microbial diversity and metabolic activity in otherwise dilute deeper oceanic layers (Fenchel and Finlay, 2008; Edgcomb, 2016). The succession of communities that occurs along oxygen gradients is well documented for prokaryotes (Noll *et al*., 2005; Chen *et al*., 2017) as well as for heterotrophic protists, such as ciliates (Fenchel and Bernard, 1996), dinoflagellates and kinetoplastids (Fenchel *et al*., 1995), and foraminiferans (Gooday *et al*., 2000).

Our results show that diplonemids and kinetoplastids are indeed sensitive to oxygen. Both groups prefer a low to medium concentration up to 7.5 mg/l and avoid areas with a higher concentration (Fig. 2 and Fig. S2). Although such a global analysis is unprecedented for both groups, several isolated reports suggested a significant presence of kinetoplastids (Edgcomb *et al*., 2011; Sauvadet *et al*., 2010) as well as diplonemids (Sauvadet *et al*., 2010) in low oxygen or anoxic habitats. In mostly bacteriovorous planktonic kinetoplastids, this phenomenon might reflect the presence of a prokaryote-rich niche. We can only speculate that the abundance of diplonemids in oxygen-poor habitats may follow the same reasoning. However, without at least partial knowledge of diplonemid biology, we cannot address this further.

On the other hand, the relatively low abundance of both groups in samples with high oxygen concentration may be caused by physiological constraints. It has been shown that protists require a much lower level of oxygen to satisfy the needs of their cellular metabolism, namely oxidative phosphorylation, compared to planktonic metazoans (Fenchel and Finlay, 1990). Also, at higher concentration, the increase in reactive oxygen species levels may slow down the growth and division of ciliates (Fenchel *et al*., 1989), so oxygen toxicity may become a significant limiting factor for both kinetoplastids and diplonemids.

Temperature has a pronounced effect on morphology, rate of metabolism and primary production of algae (Medlyn *et al*., 2002; Thomas *et al*., 2012; Toseland *et al*., 2013; Schabhüttl *et al*., 2013) as well as heterotrophic protists (Atkinson *et al*., 2003; Ikeda, 2017). Moreover, it affects the diversity and distribution patterns of protistan communities (Weisse *et al*., 2001; Schabhüttl *et al*., 2013; Gimmler *et al*., 2016; Soininen *et al*., 2016). On the global scale, this effect is described mostly as a latitudinal gradient of diversity and postulates a gradual loss of diversity from the equator to polar regions. It is well documented for large metazoans or plants (Costello and Chaudhary, 2017) but still debated for protists and marine prokaryotes (Hillebrand and Azovsky, 2001; Brayard *et al*., 2005; Fuhrman *et al*., 2008; Costello and Chaudhary, 2017; Ibarbalz *et al*., 2019; Logares *et al*., 2020). Our results suggest the existence of such a gradient for diversity and relative abundance of both diplonemids and kinetoplastids in the photic zone. This pattern is paralleled at the level of community structure as the respective analyses show clustering according to temperature and latitude in both depth zones (Figs. 3, 4, and Fig. S5).

The distribution of marine protists has been so far mostly explained by two opposing views. The first takes into account the significance of short generation times, large population sizes and high dispersal rates facilitated by the interconnection of the global oceanic ecosystem, and postulates cosmopolitan, ubiquitous distribution of most species while neglecting the effect of geographical barriers and isolation by distance (Finlay and Fenchel, 2004; Boenigk *et al*., 2012; Hellweger *et al*., 2014). The second view claims a limited dispersal and finds the level of endemism of protists comparable to other eukaryotes (Foissner, 2006). As shown recently for prokaryotes and picoeukaryotes (Logares *et al*., 2020) or diatoms (Malviya *et al*., 2016), the real-world situation probably oscillates between these two models and differs among various functional groups of plankton. Here, we show that the abundance and richness of planktonic kinetoplastids and diplonemids, as well as the structure of their communities, reflect gradients of several environmental factors. Our analyses show that the Arctic and Mediterranean samples are differentiated from the bulk of the global dataset. This may represent a response to local environmental conditions and also biogeographic patterns. However, more detailed sampling in these respective areas is necessary to identify the reasons behind such a clustering.

Methods

DNA was isolated from polycarbonate membrane filters (Isopore) using the PowerWater DNA Isolation Kit (MO BIO, USA). The V9 region of the 18S rRNA gene was amplified using universal eukaryotic primers (5'-TTGTACACACCGCCC-3′, 5'-CCTTCYGCAGGTTCACC TAC-3′, 95°C for 30 s, 62°C for 30 s, 72°C for 30 s, 25 cycles, product length 120–140 bp). V9 amplicons were sequenced using the Illumina HiSeq PE150 cycle

technology at Genome Quebec. Primer sequences were removed from reads using cutadapt v. 1.13 under the following settings: --no-indels, --discard-untrimmed, --minimum-length 50, --overlap 4, –e 0.2, –a forward primer…complement of the reverse primer, –A reverse primer…complement of the forward primer. Then reads were merged using bbmerge under the default settings. Using bbduk, merged reads containing undetermined bases (Ns) were filtered out and reads having average Phred quality below 20 were discarded. Cleaned reads were collapsed into barcodes using vsearch v. 2.4.3 under the default settings. Barcodes were grouped into OTUs using swarm v. 2.2.2 (Mahé *et al*., 2015) under the following settings: -d 1, –f, –z. For OTU definition we used barcodes from a combined pool of 57 newly reported samples (Supporting Information Table S1), and Tara Oceans V9 barcodes having an abundance of three or more reads occurring at two or more sampling locations. Barcodes from newly reported samples were taken to the OTU definition step if they were present in the Tara Oceans dataset, or if they had an abundance of three or more reads occurring at two or more sampling locations. In total, 6 173 252 barcodes were grouped into 385 783 OTUs (32 338 OTUs were present in our set of 57 samples). In total, 4 922 155 barcodes or about 80% were present in the Tara Oceans dataset only; 745 807 barcodes (12%) were shared between the Tara Oceans dataset and the set of 57 novel samples; 505 290 barcodes (8%) were present in the novel sample set only. Using ggsearch36, the OTUs were taxonomically assigned. The PR2 eukaryotic 18S rDNA database (Guillou *et al*., 2013) and an in-house 18S rDNA database of diplonemids and kinetoplastids were used as a reference for taxonomic assignment. The V9 region was extracted before the taxonomic assignment. If OTUs had an identity to a reference sequence of less than 80%, they were annotated as unknown Eukaryota. All further statistical analyses were performed on 66 870 diplonemid OTUs (2 124 OTUs were present in the set of 57 novel samples), and 544 kinetoplastid OTUs (104 OTUs in the novel sample set). As a result, OTUs were redefined and reannotated taxonomically in our dataset, as compared to the Tara Oceans dataset analysed by Flegontova *et al*. (2016, 2018).

All statistical analyses were performed using R v. 3.6.1. We calculated pairwise Spearman rank correlation coefficients for all environmental variables and applied hierarchical clustering to this matrix (the hclust R package, method 'complete'). Approximated distributions of biological variables (diplonemid richness, etc.) versus environmental variables using generalized additive models (GAM) were implemented in the mgcv R package. In most cases, outliers on the environmental variable axis that lay at a distance of more than three interquartile ranges from the first and third quartiles were removed. Then we fitted GAMs ($y \sim s(x)$) to the remaining points: GAMs based on the beta distribution that is suitable for variables between 0 and 1 (relative abundance) or GAMs based on the gamma distribution that is suitable for positive continuous variables (richness). In the latter case, the setting 'link = identity' was used.

The vegan package v. 2.5–6 with default settings for running redundancy analysis (RDA) was used, with an input matrix being read counts for diplonemid or kinetoplastid OTUs. Diplonemid OTUs with a total abundance of >9 reads were used for this analysis, while in the case of kinetoplastids, all OTUs were used. The matrix was subjected to the Hellinger transformation. RDA was performed independently for various data subsets: all data, photic zone samples, Tara Oceans mesopelagic zone samples, Arctic samples, Mediterranean samples, and samples with low oxygen concentration (from 0 to 5 mg/l). For the latter three datasets, we used just a few predictor variables that were available for the newly generated samples (depth, latitude, longitude, oxygen concentration, salinity, temperature). For the other datasets, we considered additional variables out of the 47 available. Most important predictors were chosen using either an ANOVA analysis or an ordistep analysis from the vegan R package. Then these two sets of predictors were merged, and depth, latitude, longitude, oxygen concentration, salinity, and temperature were added if they were missing in the two predictor sets.

Full-length 18S rDNA was amplified using a Phusion polymerase, a universal eukaryotic forward primer (5'-ACCTGGTTGATCCTGCCAG-3′) and a diplonemid-specific reverse primer (5'-CCCAAACRGAAGRGTYGC CCCAAC-3′): 98°C for 30 s, (98°C for 10 s, 65°C for 30 s, 72°C for 30 s) 35 cycles, 72°C for 10 min. PCR products were cloned in a TOPO vector in *E. coli* using a standard protocol and sequenced using the Sanger protocol.

Acknowledgements

The authors acknowledge support from the Czech Grant Agency, project 18-23787S (to A.H.), and the Czech Ministry of Education, Youth and Sports ERC CZ LL1601, and the ERD 0000759 project, and the Gordon and Betty Moore Foundation #9354 (to J.L.). This work was also supported by The Czech Ministry of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project "e-Infrastructure CZ – LM2018140". V.P.E. acknowledges support from OCE-1336082. Adriatic part of the study was carried out within the Croatian Science Foundation as a part of research projects: IP-2014-09-4143 'Marine microbial food web processes in global warming perspective' (MICROGLOB). We also thank the crew of R/V BIOS DVA for sampling.

Data Availability Statement

Novel metabarcoding samples presented herein are available at EBI under the following accession numbers: ERS3533235–ERS3533291. Full-length rDNA sequences are available at NCBI under accession numbers KX189173–KX189123.

References

Atkins, M.S., Teske, A.P., and Anderson, O.R. (2000) A survey of flagellate diversity at four deep-sea hydrothermal vents in the eastern Pacific Ocean using structural and molecular approaches. *J Euk Microbiol* **47**: 400–411..

Atkinson, D., Ciotti, B.J., and Montagnes, D.J.S. (2003) Protists decrease in size linearly with temperature: ca. 2.5%° C. *Proc R Soc Lond B Biol Sci* **270**: 2605–2611. https://doi.org/10.1098/rspb.2003.2538.

Boenigk, J., Ereshefsky, M., Hoef-Emden, K., Mallet, J., and Bass, D. (2012) Concepts in protistology: species definitions and boundaries. *Eur J Protistol* **48**: 96–102. https://doi.org/10.1016/j.ejop.2011.11.004.

Bråte, J., Krabberød, A.K., Dolven, J.K., Ose, R.F., Kristensen, T., Bjørklund, K.R., and Shalchian-Tabrizi, K. (2012) Radiolaria associated with large diversity of marine Alveolates. *Protist* **163**: 767–777. https://doi.org/10.1016/J.PROTIS.2012.04.004.

Brayard, A., Escarguel, G., and Bucher, H. (2005) Latitudinal gradient of taxonomic richness: combined outcome of temperature and geographic mid-domains effects? *J Zoolog Syst Evol Res* **43**: 178–188. https://doi.org/10.1111/j.1439-0469.2005.00311.x.

Brown, P.B., and Wolfe, G.V. (2006) Protist genetic diversity in the acidic hydrothermal environments of Lassen volcanic National Park, USA. *J Euk Microbiol* **53**: 420–431. https://doi.org/10.1111/j.1550-7408.2006.00125.x.

Calbet, A., and Landry, M.R. (2004) Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. *Limnol Oceanogr* **49**: 51–57. https://doi.org/10.4319/lo.2004.49.1.0051.

Callahan, H.A., Litaker, R.W., and Noga, E.J. (2002) Molecular taxonomy of the suborder Bodonina (order Kinetoplastida), including the important Fish parasite, *Ichthyobodo necator*. *J Euk Microbiol* **49**: 119–128. https://doi.org/10.1111/j.1550-7408.2002.tb00354.x.

Cavalier-Smith, T. (2016) Higher classification and phylogeny of Euglenozoa. *Eur J Protistol* **56**: 250–276. https://doi.org/10.1016/J.EJOP.2016.09.003.

Chambouvet, A., Morin, P., Marie, D., and Guillou, L. (2008) Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science* **322**: 1254–1257. https://doi.org/10.1126/science.1164387.

Chen, J., Hanke, A., Tegetmeyer, H.E., Kattelmann, I., Sharma, R., Hamann, E., *et al*. (2017) Impacts of chemical gradients on microbial community structure. *ISME J* **11**: 920–931. https://doi.org/10.1038/ismej.2016.175.

Costello, M.J., and Chaudhary, C. (2017) Marine biodiversity, biogeography, Deep-Sea gradients, and conservation. *Curr Biol* **27**: R511–R527. https://doi.org/10.1016/J.CUB.2017.04.060.

Countway, P.D., Gast, R.J., Dennett, M.R., Savai, P., Rose, J.M., and Caron, D.A. (2007) Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the western North Atlantic (Sargasso Sea and gulf stream). *Environ Microbiol* **9**: 1219–1232. https://doi.org/10.1111/j.1462-2920.2007.01243.x.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., *et al*. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**: 1261605. https://doi.org/10.1126/science.1261605.

del Campo, J., Sieracki, M.E., Molestina, R., Keeling, P., Massana, R., and Ruiz-Trillo, I. (2014) The others: our biased perspective of eukaryotic genomes. *Trends Ecol Evol* **29**: 252–259. https://doi.org/10.1016/J.TREE.2014.03.006.

Ebenezer, T., Zoltner, M., Burrel, A., Nenarokova, A., Vanclová, A., Prasad, B., *et al*. (2019) Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biol* **17**: 11.

Edgcomb, V.P. (2016) Marine protist associations and environmental impacts across trophic levels in the twilight zone and below. *Curr Opin Microbiol* **31**: 169–175. https://doi.org/10.1016/j.mib.2016.04.001.

Edgcomb, V.P., 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. https://doi.org/10.1038/ismej.2011.6.

El-Sayed, N.M., Myler, P.J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., *et al*. (2005) Comparative genomics of trypanosomatids. *Science* **309**: 404–409. https://doi.org/10.1126/science.1112181.

Elbrächter, M., Schnepf, E., and Balzer, I. (1996) *Hemistasia phaeocysticola* (Scherffel) comb. nov., Redescription of a free-living, marine, Phagotrophic Kinetoplastid flagellate. *Arch Protistenkunde* **147**: 125–136. https://doi.org/10.1016/S0003-9365(96)80028-5.

Eloe, E.A., Shulse, C.N., Fadrosh, D.W., Williamson, S.J., Allen, E.E., and Bartlett, D.H. (2011) Compositional differences in particle-associated and free-living microbial assemblages from an extreme deep-ocean environment. *Environ Microbiol Rep* **3**: 449–458. https://doi.org/10.1111/j.1758-2229.2010.00223.x.

Fenchel, T., and Bernard, C. (1996) Behavioural responses in oxygen gradients of ciliates from microbial mats. *Eur J Protistol* **32**: 55–63. https://doi.org/10.1016/S0932-4739(96)80039-3.

Fenchel, T., Bernard, C., Esteban, G., Finlay, B.J., Hansen, P.J., and Iversen, N. (1995) Microbial diversity and activity in a Danish Fjord with anoxic deep water. *Ophelia* **43**: 45–100. https://doi.org/10.1080/00785326.1995.10430576.

Fenchel, T., and Finlay, B.J. (1990) Oxygen toxicity, respiration and behavioural responses to oxygen in free-living anaerobic ciliates. *J Gen Microbiol* **136**: 1953–1959. https://doi.org/10.1099/00221287-136-10-1953.

Fenchel, T., and Finlay, B.J. (2008) Oxygen and the spatial structure of microbial communities. *Biol Rev* **83**: 553–569. https://doi.org/10.1111/j.1469-185X.2008.00054.x.

Fenchel, T., Finlay, B.J., and Giannì, A. (1989) Microaerophily in ciliates: responses of an Euplotes species

(hypotrichida) to oxygen tension. *Arch Protistenkunde* **137**: 317–330. https://doi.org/10.1016/S0003-9365(89)80015-6.

Finlay, B.J., and Fenchel, T. (2004) Cosmopolitan metapopulations of free-living microbial eukaryotes. *Protist* **155**: 237–244.

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. https://doi.org/10.1016/j.cub.2016.09.031.

Flegontova, O., Flegontov, P., Malviya, S., Poulain, J., de Vargas, C., Bowler, C., *et al*. (2018) Neobodonids are dominant kinetoplastids in the global ocean. *Environ Microbiol* **20**: 878–889. https://doi.org/10.1111/1462-2920.14034.

Foissner, W. (2006) Biogeography and dispersal of microorganisms: a review emphasizing protists. *Acta Protozool* **45**: 111–136.

Fuhrman, J.A., Steele, J.A., Hewson, I., Schwalbach, M.S., Brown, M.V., Green, J.L., and Brown, J.H. (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci U S A* **105**: 7774–7778. https://doi.org/10.1073/pnas.0803070105.

Gawryluk, R.M.R., del Campo, J., Okamoto, N., Strassert, J.F.H., Lukeš, J., Richards, T.A., *et al*. (2016) Morphological identification and single-cell genomics of marine diplonemids. *Curr Biol* **26**: 3053–3059. https://doi.org/10.1016/j.cub.2016.09.013.

Gimmler, A., Korn, R., De Vargas, C., Audic, S., and Stoeck, T. (2016) The Tara oceans voyage reveals global diversity and distribution patterns of marine planktonic ciliates. *Sci Rep* **6**: 33555. https://doi.org/10.1038/srep33555.

Giner, C.R., Forn, I., Romac, S., Logares, R., de Vargas, C., and Massana, R. (2016) Environmental sequencing provides reasonable estimates of the relative abundance of specific picoeukaryotes. *Appl Environ Microbiol* **82**: 4757–4766. https://doi.org/10.1128/AEM.00560-16.

Gooday, A.J., Bernhard, J.M., Levin, L.A., and Suhr, S.B. (2000) Foraminifera in the Arabian Sea oxygen minimum zone and other oxygen-deficient settings: taxonomic composition, diversity, and relation to metazoan faunas. *Deep Sea Res Part 2 Top Stud Oceanogr* **47**: 25–54. https://doi.org/10.1016/S0967-0645(99)00099-5.

Guillou, L., Viprey, M., Chambouvet, A., Welsh, R.M., Kirkham, A.R., Massana, R., *et al*. (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to *Syndiniales* (*Alveolata*). *Environ Microbiol* **10**: 3349–3365. https://doi.org/10.1111/j.1462-2920.2008.01731.x.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., and Christen, R. (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–D604. https://doi.org/10.1093/nar/gks1160.

Hamilton, P.B., Gibson, W.C., and Stevens, J.R. (2007) Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. *Mol Phyl Evol* **44**: 15–25. https://doi.org/10.1016/j.ympev.2007.03.023.

Hellweger, F.L., van Sebille, E., and Fredrick, N.D. (2014) Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science* **345**: 1346–1349.

Hillebrand, H., and Azovsky, A.I. (2001) Body size determines the strength of the latitudinal diversity gradient. *Ecography* **24**: 251–256. https://doi.org/10.1111/j.1600-0587.2001.tb00197.x.

Ibarbalz, F.M., Henry, N., Brandao, M.C., Martini, S., Busseni, G., Byrne, H., *et al*. (2019) Global trends in marine plankton diversity across kingdoms of life. *Cell* **179**: 1084–1097. https://doi.org/10.1016/j.cell.2019.10.008.

Ikeda, T. (2017) An analysis of metabolic characteristics of planktonic heterotrophic protozoans. *J Plankton Res* **39**: 479–490. https://doi.org/10.1093/plankt/fbx015.

Keeling, P.J., and del Campo, J. (2017) Marine protists are not just big bacteria. *Curr Biol* **27**: R541–R549. https://doi.org/10.1016/j.cub.2017.03.075.

Kelly, R.P., Shelton, A.O., and Gallego, R. (2019) Understanding PCR processes to draw meaningful conclusions from environmental DNA studies. *Sci Rep* **9**: 12133.

Kent, M.L., Elston, R.A., Nerad, T.a., and Sawyer, T.K. (1987) An *Isonema*-like flagellate (protozoa: Mastigophora) infection in larval geoduck clams, *Panope abrupta*. *J Inv Pathol* **50**: 221–229. https://doi.org/10.1016/0022-2011(87)90086-3.

Khomich, M., Kauserud, H., Logares, R., Rasconi, S., and Andersen, T. (2017) Planktonic protistan communities in lakes along a large-scale environmental gradient. *FEMS Microbiol Ecol* **93**: fiw231.

Lara, E., Moreira, D., Vereshchaka, A., and López-García, P. (2009) Pan-oceanic distribution of new highly diverse clades of deep-sea diplonemids. *Environ Microbiol* **11**: 47–55. https://doi.org/10.1111/j.1462-2920.2008.01737.x.

Larsen, J., and Patterson, J. (1990) Some flagellates (Protista) from tropical marine sediments. *J Nat Hist* **24**: 801–937.

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* **348**: 1262073. https://doi.org/10.1126/science.1262073.

Lin, X., Scranton, M., Varela, R., Chistoserdov, A., and Taylor, G. (2007) Compositional responses of bacterial communities to redox gradients and grazing in the anoxic Cariaco Basin. *Aquat Microb Ecol* **47**: 57–72. https://doi.org/10.3354/ame047057.

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., *et al*. (2014) Patterns of rare and abundant marine microbial eukaryotes. *Curr Biol* **24**: 813–821. https://doi.org/10.1016/j.cub.2014.02.050.

Logares, R., Deutschmann, I.M., Junger, P.C., Giner, C.R., Krabberød, A.K., Schmidt, T.S.B., *et al*. (2020) Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* **8**: 55. https://doi.org/10.21203/rs.2.17228/v1.

Longhurst, A.R. (2007) *Ecological Geography of the Sea*, London: Academic Press. https://doi.org/10.1016/B978-0-12-455521-1.X5000-1.

López-García, P., Philippe, H., Gail, F., and 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. https://doi.org/10.1073/pnas.0235779100.

Lukeš, J., Skalický, T., Týč, J., Votýpka, J., and Yurchenko, V. (2014) Evolution of parasitism in kinetoplastid flagellates. *Mol Biochem Parasitol* **195**: 115–122. https://doi.org/10.1016/j.molbiopara.2014.05.007.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. (2015) Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* **3**: e1420. https://doi.org/10.7717/peerj.1420.

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–E1525.

Maslov, D.A., Votýpka, J., Yurchenko, V., and Lukeš, J. (2013) Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. *Trends Parasitol* **29**: 43–52. https://doi.org/10.1016/j.pt.2012.11.001.

Massana, R., del Campo, J., Sieracki, M.E., Audic, S., and Logares, R. (2014) Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *ISME J* **8**: 854–866. https://doi.org/10.1038/ismej.2013.204.

Medlyn, B.E., Dreyer, E., Ellsworth, D., Forstreuter, M., Harley, P.C., Kirschbaum, M.U.F., *et al*. (2002) Temperature response of parameters of a biochemically based model of photosynthesis. II. A review of experimental data. *Plant Cell Environ* **25**: 1167–1179. https://doi.org/10.1046/j.1365-3040.2002.00891.x.

Moreira, D., López-García, P., and 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. https://doi.org/10.1099/ijs.0.63081-0.

Morgan-Smith, D., Herndl, G., van Aken, H., and Bochdansky, A. (2011) Abundance of eukaryotic microbes in the deep subtropical North Atlantic. *Aquat Microb Ecol* **65**: 103–115. https://doi.org/10.3354/ame01536.

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

Noll, M., Matthies, D., Frenzel, P., Derakshani, M., and Liesack, W. (2005) Succession of bacterial community structure and diversity in a paddy soil oxygen gradient. *Environ Microbiol* **7**: 382–395. https://doi.org/10.1111/j.1462-2920.2005.00700.x.

O'Neill, E.C., Trick, M., Henrissat, B., and Field, R.A. (2015) Euglena in time: evolution, control of central metabolic processes and multi-domain proteins in carbohydrate and natural product biochemistry. *Perspect Sci* **6**: 84–93. https://doi.org/10.1016/J.PISC.2015.07.002.

Okamoto, N., Gawryluk, R.M.R., del Campo, J., Strassert, J. F.H., Lukeš, J., Richards, T.A., *et al*. (2019) A revised taxonomy of Diplonemids including the Eupelagonemidae n. fam. and a type species, *Eupelagonema oceanica* n. gen. and n. sp. *J Euk Microbiol* **66**: 519–524. https://doi.org/10.1111/jeu.12679.

Pachiadaki, M.G., Taylor, C., Oikonomou, A., Yakimov, M. M., Stoeck, T., and Edgcomb, V.P. (2016) In situ grazing experiments apply new technology to gain insights into deep-sea microbial food webs. *Deep Sea Res Part 2 Top Stud Oceanogr* **129**: 223–231. https://doi.org/10.1016/J.DSR2.2014.10.019.

Pernice, M.C., Giner, C.R., Logares, R., Perera-Bel, J., Acinas, S.G., Duarte, C.M., *et al*. (2015) Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J* **10**: 1–14. https://doi.org/10.1038/ismej.2015.170.

Porter, D. (1973) *Isonema papillatum* sp. n., a new colorless marine flagellate: a light- and electronmicroscopic study. *J Protozool* **20**: 351–356. https://doi.org/10.1111/j.1550-7408.1973.tb00895.x.

Righetti, D., Vogt, M., Gruber, N., Psomas, A., and Zimmermann, N.E. (2019) Global pattern of phytoplankton diversity driven by temperature and environmental variability. *Sci Adv* **5**: eaau6253.

Rocke, E., Pachiadaki, M.G., Cobban, A., Kujawinski, E.B., and Edgcomb, V.P. (2015) Protist community grazing on prokaryotic prey in deep ocean water masses. *PLoS One* **10**: e0124505. https://doi.org/10.1371/journal.pone.0124505.

Rodríguez-Zavala, J.S., Ortiz-Cruz, M.A., Mendoza-Hernández, G., and Moreno-Sánchez, R. (2010) Increased synthesis of α-tocopherol, paramylon and tyrosine by *Euglena gracilis* under conditions of high biomass production. *J Appl Microbiol* **109**: 2160–2172. https://doi.org/10.1111/j.1365-2672.2010.04848.x.

Salani, F.S., Arndt, H., Hausmann, K., Nitsche, F., and Scheckenbach, F. (2012) Analysis of the community structure of abyssal kinetoplastids revealed similar communities at larger spatial scales. *ISME J* **6**: 713–723. https://doi.org/10.1038/ismej.2011.138.

Sauvadet, A.L., Gobet, A., and Guillou, L. (2010) Comparative analysis between protist communities from the deep-sea pelagic ecosystem and specific deep hydrothermal habitats. *Environ Microbiol* **12**: 2946–2964. https://doi.org/10.1111/j.1462-2920.2010.02272.x.

Schabhüttl, S., Hingsamer, P., Weigelhofer, G., Hein, T., Weigert, A., and Striebel, M. (2013) Temperature and species richness effects in phytoplankton communities. *Oecologia* **171**: 527–536. https://doi.org/10.1007/s00442-012-2419-4.

Scheckenbach, F., Hausmann, K., Wylezich, C., Weitere, M., and 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. https://doi.org/10.1073/pnas.0908816106.

Šimek, K., Hartman, P., Nedoma, J., Pernthaler, J., Springmann, D., Vrba, J., and Psenner, R. (1997) Community structure, picoplankton grazing and zooplankton control of heterotrophic nanoflagellates in a eutrophic reservoir during the summer phytoplankton maximum. *Aquat Microb Ecol* **12**: 49–63. https://doi.org/10.3354/ame012049.

Simpson, A.G., and Roger, A.J. (2004) Protein phylogenies robustly resolve the deep-level relationships within Euglenozoa. *Mol Phyl Evol* **30**: 201–212. https://doi.org/10.1016/S1055-7903(03)00177-5.

Simpson, A.G.B., Lukeš, J., and Roger, A.J. (2002) The evolutionary history of Kinetoplastids and their Kinetoplasts.

Mol Biol Evol **19**: 2071–2083. https://doi.org/10.1093/oxfordjournals.molbev.a004032.

Soininen, J., Jamoneau, A., Rosebery, J., and Passy, S.I. (2016) Global patterns and drivers of species and trait composition in diatoms. *Glob Ecol Biogeogr* **25**: 940–950. https://doi.org/10.1111/geb.12452.

Šolić, M., and Krstulović, N. (1994) Role of predation in controlling bacterial and heterotrophic nanoflagellate standing stocks in the coastal Adriatic Sea: seasonal patterns. *Mar Ecol Prog Ser* **114**: 219–235.

Šolić, M., Šantić, D., Šestanović, S., Bojanić, N., Ordulj, M., Jozić, S., and Vrdoljak, A. (2018) The effect of temperature increase on microbial carbon fluxes in the Adriatic Sea: an experimental approach. *FEMS Microbiol Ecol* **94**: fiy169.

Šolić, M., Šantić, D., Šestanović, S., Bojanić, N., Jozić, S., Vrdoljak, A., *et al*. (2019) Temperature and phosphorus interacts in controlling the picoplankton carbon flux in the Adriatic Sea: an experimental versus field study. *Environ Microbiol* **21**: 2469–2484. https://doi.org/10.1111/1462-2920.14634.

Stock, A., Breiner, H.-W., Pachiadaki, M., Edgcomb, V.P., Filker, S., La Cono, V., *et al*. (2012) Microbial eukaryote life in the new hypersaline deep-sea basin Thetis. *Extremophiles* **16**: 21–34. https://doi.org/10.1007/s00792-011-0401-4.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.-W., and Richards, T.A. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* **19**: 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x.

Tashyreva, D., Prokopchuk, G., Yabuki, A., Kaur, B., Faktorová, D., Votýpka, J., *et al*. (2018) Phylogeny and morphology of new diplonemids from Japan. *Protist* **169**: 158–179. https://doi.org/10.1016/J.PROTIS.2018.02.001.

Thomas, M.K., Kremer, C.T., Klausmeier, C.A., and Litchman, E. (2012) A global pattern of thermal adaptation in marine phytoplankton. *Science* **338**: 1085–1088. https://doi.org/10.1126/science.1224836.

Tillmann, U. (2004) Interactions between planktonic microalgae and protozoan Grazers1. *J Euk Microbiol* **51**: 156–168. https://doi.org/10.1111/j.1550-7408.2004.tb00540.x.

Toseland, A., Daines, S.J., Clark, J.R., Kirkham, A., Strauss, J., Uhlig, C., *et al*. (2013) The impact of temperature on marine phytoplankton resource allocation and metabolism. *Nat Clim Chang* **3**: 979–984. https://doi.org/10.1038/nclimate1989.

von der Heyden, S., and 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**: 2605–2621. https://doi.org/10.1099/ijs.0.63606-0.

von der Heyden, S., Chao, E.E., Vickerman, K., and Cavalier-Smith, T. (2004) Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of euglenozoa. *J Euk Microbiol* **51**: 402–416. https://doi.org/10.1111/j.1550-7408.2004.tb00387.x.

Weisse, T., Karstens, N., Meyer, V., Janke, L., Lettner, S., and Teichgräber, K. (2001) Niche separation in common prostome freshwater ciliates: the effect of food and temperature. *Aquat Microb Ecol* **26**: 167–179. https://doi.org/10.3354/ame026167.

Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E., and Keeling, P.J. (2015) Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347**: 1257594. https://doi.org/10.1126/science.1257594.

Yabuki, A., and Tame, A. (2015) Phylogeny and reclassification of *Hemistasia phaeocysticola* (Scherffel) Elbrächter & Schnepf, 1996. *J Euk Microbiol* **62**: 426–429. https://doi.org/10.1111/jeu.12191.

Yazaki, E., Ishikawa, S.A., Kume, K., Kumagai, A., Kamaishi, T., Tanifuji, G., *et al*. (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. https://doi.org/10.1266/ggs.16-00056.

Zhao, F., Filker, S., Xu, K., Huang, P., and Zheng, S. (2017) Patterns and drivers of vertical distribution of the ciliate community from the surface to the abyssopelagic zone in the Western Pacific Ocean. *Front Microbiol* **8**: 2559. https://doi.org/10.3389/fmicb.2017.02559.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Overview and basic characteristics of samples analysed in this study. The following information is provided: sample ID, sampling location ID, sampling date, dataset name, depth zone (SRF, surface; DCM, deep chlorophyll maximum; MES, mesopelagic; OMZ, oxygen minimum zone), latitude zone, biome, ocean, oceanographic province (Longhurst, 20072007), depth (m), latitude and absolute latitude (decimal format), longitude (decimal format), salinity, temperature (°C), water density expressed as σ_θ (the density calculated with *in situ* salinity, potential temperature, and zero pressure – 1000 kg m^{-3}), oxygen concentration (mg L^{-1}), fluorescence intensity, distance to coast (km).

Fig. S1. Boxplots of diplonemid (**A**, **C**) and kinetoplastid (**B**, **D**) relative abundance (**A**, **B**) and richness (**C**, **D**) broken into datasets and depth zones. The ends of the boxes are the upper and lower quartiles. The median is marked by a bold line inside boxes. Lines outside the box mark 1.5 interquartile range.

Fig. S2. Scatterplots showing the distribution of kinetoplastid richness across six environmental variables. Samples are represented by dots coloured according to the dataset and depth zone. Blue lines represent the GAM smooth trends and grey ribbons, the corresponding 95% confidence intervals of kinetoplastid richness predicted by the GAMs. If the p-value of smooth terms is <0.05, the deviance explained is shown in the upper right corner of the plot. Otherwise, it is written 'n.-s.' (nonsignificant) in the upper right corner of the plot. (**A**) Scatterplots showing the distribution of kinetoplastid richness for the combined dataset. (**B**) Scatterplots showing the distribution of kinetoplastid richness for only the Tara dataset. (**C**) Dendrogram generated by hierarchical

clustering analysis shows pairwise Spearman rank correlation coefficients for seven environmental variables.

Fig. S3. Scatterplots showing the distribution of four summary statistics, diplonemid and kinetoplastid richness and relative abundance across 23 environmental variables for Tara dataset. Samples are represented by dots coloured according to depth zone (cyan is photic, pink is mesopelagic). Blue lines represent the GAM smooth trends and grey ribbons, the corresponding 95% confidence intervals of variables predicted by the GAMs. If the p-value of smooth terms is <0.05, the deviance explained is shown in the upper right corner of the plot. Otherwise, it is written 'n.-s.' (nonsignificant) in the upper right corner of the plot.

Fig. S4. Dendrogram generated by hierarchical clustering analysis shows pairwise Spearman rank correlation coefficients for 47 environmental variables across combined dataset (**A**); for 47 environmental variables across photic Tara, Arctic, Adriatic, and Cariaco samples (**B**); for 39 environmental variables across mesopelagic Tara, Arctic, Adriatic, and Cariaco samples (**C**); for seven environmental variables across Arctic dataset (**D**); for six environmental variables across our Adriatic and Tara Mediterranean samples (**E**); for six environmental variables across our Cariaco and Tara samples with oxygen concentration below 5 mg/l (**F**).

Fig. S5. Impact of environmental variables on the kinetoplastid community structure from the combined dataset. The redundancy analysis (RDA) plot computed on 539 kinetoplastid swarms from 287 samples, using eighteen environmental predictor variables after selection and two biological predictor variables (richness and relative abundance of kinetoplastids). The adjusted R-squared of the analysis is of 0.248745 (0.3012803 unadjusted). Samples are represented by dots coloured according to the dataset and depth zone. Arrows correspond to the predictor variables.

Fig. S6. Impact of environmental variables on the community structure of kinetoplastid from photic (**A**) and mesopelagic (**B**) depth zones. (**A**) The redundancy analysis (RDA) plot computed on 488 kinetoplastid swarms from 201 photic samples, using nineteen environmental predictor variables after selection and two biological predictor variables (richness and relative abundance of kinetoplastids). The adjusted R-squared of the analysis is of 0.1854501 (0.2709778 unadjusted). Samples are represented by dots coloured according to geographical regions and depth zone for non-polar samples. Arrows correspond to the predictor variables. (**B**) The redundancy analysis (RDA) plot computed on 233 kinetoplastid swarms from 35 Tara mesopelagic samples, using eleven environmental predictor variables after selection and two biological predictor variables (richness and relative abundance of kinetoplastids). The adjusted R-squared of the analysis is of 0.2191754 (0.517726 unadjusted). Samples are represented by dots coloured according to biogeographical provinces. Arrows correspond to the predictor variables.

Fig. S7. Impact of environmental variables on the community structure of kinetoplastid from Arctic dataset (**A**), from Mediterranean samples (**B**), and samples with oxygen concentration below 5 mg/l (**C**). (**A**) The redundancy analysis (RDA) plot computed on 78 kinetoplastid swarms from 37 Arctic samples, using six environmental and two biological (richness and relative abundance of kinetoplastids) predictor variables. The adjusted R-squared of the analysis is of 0.3059302 (0.4601679 unadjusted). Samples are represented by dots coloured according to latitude and depth. Arrows correspond to the predictor variables. (**B**) The redundancy analysis (RDA) plot computed on 152 kinetoplastid swarms from fifteen new Adriatic samples and 34 Tara Oceans Mediterranean samples (separated by the bold line), using six environmental and two biological (richness and relative abundance of kinetoplastids) predictor variables. The adjusted R-squared of the analysis is of 0.3032702 (0.4193918 unadjusted). Samples are represented by dots coloured according to longitude. Two or more samples coming from a single sampling site are connected by a line. Arrows correspond to the predictor variables. (**C**) The redundancy analysis (RDA) plot computed on 223 kinetoplastid swarms from 30 samples with oxygen concentration below 5 mg/l, using six environmental and two biological (richness and relative abundance of kinetoplastids) predictor variables. The adjusted R-squared of the analysis is of 0.2901564 (0.4859754 unadjusted). Samples are represented by dots coloured according to depth zones and oxygen concentration for mesopelagic samples. Arrows correspond to the predictor variables.