



HAL
open science

Environmental drivers of plankton protist communities along latitudinal and vertical gradients in the oldest and deepest freshwater lake

Gwendoline M David, David Moreira, Guillaume Reboul, Nataliia V Annenkova, Luis J Galindo, Paola Bertolino, Ana I López-Archilla, Ludwig Jardillier, Purificación López-García

► To cite this version:

Gwendoline M David, David Moreira, Guillaume Reboul, Nataliia V Annenkova, Luis J Galindo, et al.. Environmental drivers of plankton protist communities along latitudinal and vertical gradients in the oldest and deepest freshwater lake. *Environmental Microbiology*, 2021, 23 (3), pp.1436-1451. 10.1111/1462-2920.15346 . hal-03043428

HAL Id: hal-03043428

<https://hal.science/hal-03043428v1>

Submitted on 7 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Environmental drivers of plankton protist communities along latitudinal and vertical gradients in the oldest and deepest freshwater lake

5

Gwendoline M. David¹, David Moreira¹, Guillaume Reboul¹, Nataliia V. Annenkova², Luis J. Galindo¹, Paola Bertolino¹, Ana I. López-Archilla³, Ludwig Jardillier¹ and Purificación López-García¹

¹ Ecologie Systématique Evolution, Centre National de la Recherche Scientifique - CNRS, Université Paris-Saclay, AgroParisTech, Orsay, France

10

² Limnological Institute, Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia

³ Departamento de Ecología, Universidad Autónoma de Madrid, Madrid, Spain

For correspondence: puri.lopez@u-psud.fr

15

Running title: Drivers of protist communities in Lake Baikal

20

Originality and Significance Statement

Lake Baikal is the oldest, deepest and most voluminous freshwater lake on Earth, offering a unique opportunity to test the effects of horizontal versus vertical gradients on microbial community structure.

25 Using a metabarcoding approach, we studied planktonic microbial eukaryotes from Baikal water columns (5 up to 1,400 m depth) across a North-South latitudinal gradient (~600 km), including coastal and pelagic areas. Our results show that depth has a strong effect on protist community assemblage, but not latitude (minor effect) or coastal vs. open water sites (no effect). Co-occurrence analyses also point to specific biotic interactions as drivers of community structure. This comprehensive survey
30 constitutes a useful reference for monitoring active climate change effects in this ancient lake.

Summary

35

Identifying which abiotic and biotic factors determine microbial community assembly is crucial to understand ecological processes and predict how communities will respond to environmental change. While global surveys aim at addressing this question in the world's oceans, equivalent studies in large freshwater systems are virtually lacking. Being the oldest, deepest and most voluminous freshwater lake on Earth, Lake Baikal offers a unique opportunity to test the effect of horizontal versus vertical gradients in community structure. Here, we characterized the structure of planktonic microbial eukaryotic communities (0.2-30 μm cell size) along a North-South latitudinal gradient (~600 km) from samples collected in coastal and pelagic waters and from surface to the deepest zones (5-1,400 m) using an 18S rRNA gene metabarcoding approach. Our results show complex and diverse protist communities dominated by alveolates (ciliates and dinoflagellates), ochrophytes and holomycotan lineages, with cryptophytes, haptophytes, katablepharids and telonemids in moderate abundance and many low-frequency lineages, including several typical marine members, such as diplomonads, syndinians and radiolarians. Depth had a strong significant effect on protist community stratification. By contrast, the effect of the latitudinal gradient was marginal and no significant difference was observed between coastal and surface open water communities. Co-occurrence network analyses showed that epipelagic communities were significantly more interconnected than deeper communities, suggesting specific biotic interactions between autotrophic, heterotrophic and parasitic lineages that influence protist community structure. Since climate change is rapidly affecting Siberia and Lake Baikal, our comprehensive protist survey constitutes a useful reference to monitor ongoing community shifts.

55

Keywords: Lake Baikal; protist; 18S rRNA gene metabarcoding; marine-freshwater transition; light stratification; network analysis

60

Introduction

Of all ecosystems, freshwater reservoirs are the most dynamic and concentrate a high biodiversity (Rolls et al., 2018). Freshwater ecosystems are particularly vulnerable to climate change owing to a higher exposure and sensitivity to increasing temperature and other altered conditions, limited dispersal across these fragmented habitats and little-known, but likely modest, resilience potential (Woodward et al., 2010; Markovic et al., 2017). Since microorganisms are crucial in biogeochemical cycles, the impact of climate change will strongly depend on how they will respond to environmental challenge (Cavicchioli et al., 2019). Permafrost-covered areas in the Arctic region (Schuur et al., 2015) and forest-steppe ecotones in Siberia are among the most heavily impacted regions by global warming (Mackay et al., 2017). This includes Lake Baikal, in southern Siberia, which is the oldest (ca. 30 Myr), deepest, and most capacious freshwater lake on Earth (Müller et al., 2001). Lake Baikal is rapidly changing, as can be told from trends in hydrological and hydrochemical processes (Moore et al., 2009; Shimaraev and Domysheva, 2012). The lake sediments represent a continuous record of past climate for over 12 million years (Kashiwaya et al., 2001; Prokopenko et al., 2002) such that Lake Baikal is a unique model to understand and predict microbial community change and how this is linked to carbon cycling and hydrological processes.

A mandatory prerequisite for such a task is to have comprehensive information about the existing microbial community structure. However, if the broad biodiversity of Lake Baikal metazoans, including many endemisms (1,455 out of 2,595 species described), has been amply documented in the past two centuries, that of microbial life is highly fragmentary. One of the reasons relates to the large dimensions of the lake, which is around 640 km long, attains a depth of ca. 1,650 m and contains around 20% of the Earth's unfrozen freshwater (Sherstyankin et al., 2006; UNDP-GEF, 2015). This, together with its geographical location and its association to a rifting zone make Lake Baikal unique and listed as UNESCO World Heritage Site (UNDP-GEF, 2015). The lake is divided in three major basins (Northern, Central, Southern) by, respectively, the Academician Ridge and the Selenga river delta (Mats and Perepelova, 2011). Its surface freezes in winter for several months, favoring coastal downwelling and deep-water oxygenation (Schmid et al., 2008; Moore et al., 2009). As a result, Lake Baikal ultra-oligotrophic waters are globally cold (~4°C) and oxygen-rich down to the bottom (Schmid et al., 2008; Moore et al., 2009; Shimaraev and Domysheva, 2012; Troitskaya et al., 2015). Baikal also uniquely hosts methane hydrates, which are stabilized by the low temperatures and high pressures (De Batist et al., 2002; Granin et al., 2019). All these features make Lake Baikal akin a freshwater sea.

Microbial diversity in Lake Baikal plankton was first studied by classical observation and cultural approaches (Maksimova and Maksimov, 1972; Maksimov et al., 2002; Bel'kova et al., 2003) before molecular tools started to be applied at the beginning of the century (Glöckner et al., 2000) and

expanded more recently with the generalization of high-throughput sequencing. Several 16S rRNA gene-based metabarcoding studies have targeted pelagic bacteria diversity (Kurilkina et al., 2016; Belikov et al., 2019; Zakharenko et al., 2019; Wilburn et al., 2020) and, more recently, metagenomic analyses have been used to characterize planktonic prokaryotic communities from sub-ice (Cabello-Yeves et al., 2018) and deep waters (Cabello-Yeves et al., 2020), virus-bacteria assemblages in coastal waters (Butina et al., 2019) or viruses from the pelagic zone (Potapov et al., 2019). Microbial eukaryotes have only been partially studied by 18S rRNA gene metabarcoding. Several of these studies focused on phytoplankton, either on specific groups, such as diatoms (Zakharova et al., 2013) or dinoflagellates (Annenkova et al., 2011), or on whole communities, from winter sub-ice waters (Bashenkhaeva et al., 2015) to spring blooms (Mikhailov et al., 2015; Mikhailov et al., 2019b; Mikhailov et al., 2019a). Remarkably few studies have aimed at charactering the diversity of all microbial eukaryotes, especially in a comparative manner. Yi et al. (2017) analyzed protist diversity by 454 sequencing of 18S rRNA gene V9-region amplicons along the Southern basin water column (52-1450 m). Mikhailov et al. (2019a) also 454-sequenced 18S rRNA gene V3-region amplicons from several surface (0-25 m) sites across basins and built co-occurrence networks including prokaryotes to explore potential interactions during a spring phytoplankton bloom. More recently, Annenkova et al. (2020) determined the community structure of small protists (0.45-8 μm cell-size fraction) from surface waters (1-15-50 m) across the lake via 18S rRNA gene V4-region metabarcoding using higher throughput sequencing approaches (Illumina MiSeq) and suggested that some clades within known protist groups might be endemic. Nonetheless, we still lack a comprehensive view about how microbial eukaryotes distribute in the lake plankton, across basins and throughout the complete water column and, crucially, which are the most influential parameters determining community structure.

In this work, we carry out a wide-ranging comparative study of Lake Baikal planktonic protist communities in the 0.2-30 μm cell-size range using a 18S rRNA gene metabarcoding approach to study distribution patterns and to test whether depth, latitude or the coastal versus pelagic location determine community structure. With this aim, we analyze 65 samples from 17 sites across a ~600 km latitudinal North-South transect along the three lake basins and from littoral shallow areas to deep water columns covering the epi-, meso- and bathypelagic region. Our results show complex and diverse protist communities that are mostly structured by depth and that include several typical marine lineages in low abundance. Network analyses show that epipelagic communities are more interconnected than meso- and bathypelagic communities, suggesting potential specific biotic interactions between autotrophs, heterotrophs and parasites.

130 Experimental procedures

Sample collection

Lake Baikal water samples were collected at different depths from seventeen sites distributed along a North-South transect during a French-Russian research cruise in the summer of 2017. Sites were chosen to cover littoral (8) and open water (9) samples, including the deepest zones in the three major basins of the lake. In total, 65 water samples were collected from depths ranging from 5 to 1,400 m and covering three depth zones in the water column (epipelagic, 0-200 m; mesopelagic, 200-1,000 m; bathypelagic, >1,000 m); deep samples were collected far from the bottom to avoid sediment disturbance (Table S1). Given the high latitude of Lake Baikal, with a concomitant more oblique incidence of sunlight, samples collected at 200 m depth were included in the mesopelagic zone. Samples were collected with Niskin bottles (5 l for epipelagic waters, 10 l for meso- and bathypelagic waters). The physicochemical parameters of lake waters were measured with a multiparameter probe Multi 350i (WTW, Weilheim, Germany). The water was sequentially filtered onboard immediately after collection through 30- μ m and 0.22- μ m pore-size Nucleopore filters (Whatman, Maidstone, UK) and 0.2 μ m pore-size Cell-Trap units (MEM-TEQ Ventures Ltd, Wigan, UK). Volumes of water samples filtered through Cell-Traps were smaller (samples indicated with an asterisk in Fig.1). The recovered biomass and biomass-containing filters were fixed in absolute ethanol and stored at -20°C until processed.

DNA purification, 18S rRNA gene-fragment amplification and sequencing

DNA was purified using the Power Soil™ DNA purification kit (Qiagen, Hilden, Germany). 18S rRNA gene fragments (~530 bp) encompassing the V4 region were PCR-amplified using EK-565F-NGS (5'-GCAGTTAAAAAGCTCGTAGT-3') and UNonMet (5'-TTTAAGTTTCAGCCTTGCG-3'), the latter biased against metazoans (Bower et al., 2004). Primers were tagged with specific 10-bp molecular identifiers (MIDs) for multiplexed sequencing. To minimize PCR-associated biases, five PCR reaction products per sample were pooled. PCR reactions were conducted in 25- μ l reaction mixtures containing 0.5-3 μ l of eluted DNA, 1.5 mM MgCl₂, 0.2 mM dNTPs, 0.3 μ M primers and 0.5 U Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA) for 35 cycles (94°C for 30 s, 55-58°C for 30-45 s, 72°C for 90 s) preceded by 2 min denaturation at 94°C and followed by 5 min extension at 72°C. Pooled amplicons were purified using QIAquick PCR purification kit (Qiagen, Hilden, Germany). Amplicons were sequenced using paired-end (2x300 bp) Illumina MiSeq (Eurofins Genomics, Ebersberg, Germany). Sequences have been deposited in GenBank under the BioProject number PRJNA657482 (BioSamples SAMN15830589 to SAMN15830657).

Sequence and phylogenetic analyses

We used an in-house bioinformatic pipeline to process raw sequences. Paired-end reads were merged with FLASH (Magoc and Salzberg, 2011) under strict criteria and assigned to specific samples based on their MID. MID and primer sequences were trimmed using CUTADAPT (Martin, 2011). Cleaned merged reads were next dereplicated to unique sequences using VSEARCH (Rognes et al., 2016), which was also used to detect and eliminate potential chimeras. Non-chimeric sequences from all samples were pooled together to define operational taxonomic units (OTUs) at a conservative threshold of 95% identity for 18S rRNA genes using CD-HIT-EST (Fu et al., 2012) and SWARM (Mahe et al., 2015). Singletons were excluded from subsequent analyses. OTUs were assigned to taxa based on their similarity with a local 18S rRNA database build from SILVA v128 (Quast et al., 2013) and PR2 v4.5 (Guillou et al., 2013). OTUs less than 80% identical to their best environmental hit were blasted against the GenBank *nr* database (<https://ncbi.nlm.nih.gov/>) and assigned manually by phylogenetic placement analyses. Briefly, the closest hits to our OTUs in SILVA and PR2 were aligned with full 18S rDNA reference sequences covering the eukaryotic diversity using MAFFT (Katoh and Standley, 2013). After removal of uninformative sites with trimAl (Capella-Gutierrez et al., 2009), we built a tree with full reference sequences with IQ-tree (Nguyen et al., 2015) under a GTR+G+I sequence evolution model. OTU sequences were aligned to the reference alignment and then placed in the reference phylogenetic tree using EPA-ng (Barbera et al., 2019). OTUs with no reliable affiliation were maintained as 'Unclassified'. Maximum likelihood phylogenetic trees of diplomid and radiolarian OTUs were reconstructed from specific MAFFT alignments including their closest blast hits and reference sequences with PhyML (Guindon and Gascuel, 2003) applying a GTR+G+I (4 categories) model of sequence evolution. Bootstrap values were obtained from 100 replicates.

185

Statistical analyses

We generated a table of eukaryotic OTU read abundance in the different samples of Lake Baikal for diversity and statistical analyses (Tables S1-S2). To avoid biases due to differences in absolute numbers of reads per sample, we rarefied our sequences to the second smaller number of reads (9771 in BK16.500m). BK28.100m was excluded from this process due to its lower number of reads. Statistical analyses were conducted on these data with R (R Development Core Team, 2017). Richness and diversity indices were calculated using the vegan package (Oksanen et al., 2011). Evenness was calculated according to Pielou (Pielou, 1966). To see if these indices were significantly different between sampling depths and basins, we performed Wilcoxon tests between the groups distributions using R. Likewise, to test the effect of sampling point, basin and depth class on protist community composition across samples, we conducted permutational multivariate analyses of variance (PERMANOVA) based on Wisconsin-standardized Bray-Curtis dissimilarities, using the adonis function

195

of the vegan package. Across-sample community composition differences were visualized using non-metric multidimensional scaling (NMDS) analysis, also on Wisconsin-standardized Bray-Curtis dissimilarities. To connect communities according to specific origin we drew ellipses with the ade4 package (Dray and Dufoour, 2007). To test the significance of groups revealed by NMDS, we applied analysis of similarity (ANOSIM) tests with 999 permutations. Principal component analysis (PCA) of abiotic parameters based on centered and scaled data was performed with FactoMineR (Lê et al., 2008).

205 Network analysis

We built co-occurrence networks for each depth category (epipelagic, mesopelagic, bathypelagic) and a combination of meso- and bathypelagic samples (dark water column) using a multivariate Poisson lognormal model with the R-package PLNmodels (Chiquet et al., 2018) in order to account for depth-class differences between samples and potential additional covariables (specifically the sampling basin). We retained for the analysis OTUs present in more than 20% of samples and abundances higher than 0.01%. For model selection, we used Bayesian information criteria with a 50-size grid of penalties. Networks were visualized with the ggnet R-package (Chiquet et al., 2018). To further analyze network structure, we carried out a block model analysis using a stochastic block model approach on the binary co-occurrence network using the blockmodel R-package (Leger, 2016), which synthetizes the overall network structure by gathering nodes in groups with similar modes of interactions. Network properties were calculated using the igraph R-package (Csardi and Nepusz, 2006). Properties included the number of positive and negative edges, the total number of nodes and number of connected nodes. Network mean degrees correspond to the average number of established edges. The average path length indicates the mean number of edges necessary to link a given node randomly to another. Network complexity was estimated using two indicators: connectance and clustering coefficient. The connectance was calculated as $c = \frac{2E}{N \times (N-1)}$, where E is the number of edges and N the number of nodes (Barrat et al., 2008). Connectance is 1 when all possible links are established. The clustering coefficient is the probability that two nodes having a similar neighbor are connected to each other (Delmas et al., 2019). It varies between 0 and 1; low values indicated poor connectivity. Since the number of samples was different for the three depths classes, a bootstrap analysis (2,500 randomizations) was performed to test for the significance of the differences observed between the four networks. For instance, to compare the properties of the epipelagic and bathypelagic networks, 7 samples were selected randomly 2,500 times among the 42 epipelagic samples. The distribution of the randomized epipelagic network properties was then compared to that of the bathypelagic network.

230

Results and discussion

Abiotic variables across sampling sites

235 We collected Lake Baikal water samples along the Northern, Central and Southern basins from the same established depths in the water column (except for the deepest sample, which was collected close to the bottom but at sufficient distance –minimum 45 m– to avoid sediment influence) (Fig.1A; Table S1). Samples from coastal areas were always collected at 5 m depth in the water column. The measured physicochemical parameters were remarkably stable across sites and depths. Temperature ranged
240 from 3.6 to 15.3°C, but was globally low (average 5.7°C; only five surface samples exceeded 9.5°C), and significantly higher in epipelagic samples (Fig.S1). pH ranged from 7.45 to 8.47. Salinity was extremely low (always 0.0 PSU) as, accordingly, conductivity and total dissolved solids (TDS). Dissolved oxygen was high (mean 79.5%). Like temperature, pH, conductivity and dissolved oxygen in mesopelagic waters were significantly lower than in epipelagic samples. Bathypelagic parameters were similar to those of
245 the mesopelagic zone but more variable. In terms of basins, temperature, pH, conductivity and dissolved oxygen were higher in the Southern basin, which is also more impacted by human activities and pollution, notably aromatic hydrocarbons and mercury brought by the Selenga river (Adams et al., 2018; Roberts et al., 2020), although only oxygen and, marginally, conductivity were significantly different (Fig.S1). The two main axes of a PCA considering these abiotic parameters explained 58% of
250 the variance (Fig.1B). Surface samples correlated with higher temperature, conductivity and, to a lower extent, pH and dissolved oxygen. These observations suggest that depth, as a proxy for light accessibility but also temperature and other abiotic parameters, might be a strong environmental driver for community structure.

255 Composition of planktonic protist communities

To study the diversity and relative abundance of microbial eukaryotes in Lake Baikal plankton, we concentrated cells in the 0.2-30 µm diameter fraction by successive filtration steps. This fraction thus integrated pico- (0.2-2 µm), nano- (2-20 µm) and small microplankton (20-30 µm), covering a wider protistan spectrum than some previous comparative studies (Annenkova et al., 2020). We purified DNA
260 and massively sequenced (MiSeq Illumina, 2x300 bp) multiplexed 18S rRNA gene V4-region amplicons. After discarding low-quality reads, we generated 6,405,343 high-quality merged paired-end sequences that we clustered in operational taxonomic units (OTUs) at different thresholds. We determined 27,504 OTUs and 9,700 OTUs at, respectively, 98% and 95% sequence identity (CD-HIT). SWARM yielded 11,590 OTUs (Table S1), only slightly higher than the number of OTUs defined at the latter cut-off. For
265 subsequent comparative analyses, we used OTUs derived both from SWARM and defined at 95%

sequence identity threshold, which yielded very similar results (see below). Many diversity studies focus on exact sequence variants after sequence error correction (Callahan et al., 2016) that can inform about individual strain variation. However, for the purpose of this comparative study, we chose to use conservatively defined OTUs that, on average (this varies across phylogenetic groups), correspond to the genus or species-genus level (Caron et al., 2009). This taxonomy cut-off level is relevant for broad comparative ecological studies (members of the same genus are likely to have similar general functions, despite inter- strain or species-specific niche differences), while operationally diminishing the number of handled OTUs. In addition, due to the SWARM iterative growth process and the use of sequence abundance to delineate OTUs, SWARM can distinguish OTUs different in as little as 2 positions between their centres, which is a precision almost comparable to that of amplicon sequence variants. Based on sequence MIDs, the abundance of the different OTUs was determined for each sample (Table S1). To avoid potential biases in diversity and relative abundance estimates linked to differences in the total number of reads, we rarefied sequences to the same number across samples, which resulted in a global number of 4,570 genus-level OTUs. Nonetheless, accumulation curves showed that the diversity of planktonic protists was far from reaching saturation, even at the conservative genus level (Fig.S2). Richness significantly decreased in deep as compared to surface waters; so did evenness (Fig.S3). We did not observe richness differences across lake basins, but evenness appeared significantly higher in the Northern basin.

From a phylogenetic perspective, our defined OTUs affiliated to at least 27 eukaryotic phyla belonging to several major eukaryotic supergroups (Fig.1C; Table S2): the SAR clade (Stramenopiles, Alveolata, Rhizaria), Amoebozoa, Archaeplastida, Excavata, Opisthokonta and Hacrobia. Although we considered Hacrobia as originally described (Okamoto et al., 2009), they should be possibly split in two or more groups as the eukaryotic phylogeny progressively resolves (Burki et al., 2020). Ciliates and dinoflagellates (Alveolata), Ochrophyta (Stramenopiles) and Holomycota (Fungi and related lineages within the Opisthokonta) dominated plankton samples representing, respectively 48.4%, 21.5%, 12.6% and 8% relative sequence abundance. Cryptophyta, Haptophyta, Kathablepharida and Tenomemida displayed moderate abundances (0.5 to 5% reads) and were followed by a long tail of lower-frequency taxa in rank:abundance curves (Fig.S4). The major dominant groups were similar in all depths, with small variations in the deepest waters. However, at lower taxonomic levels (i.e. OTU level), there were important differences. Indeed, microbial eukaryotic composition was specific to each depth level (Fig. 3C). Only 7.44% of the rarefied OTUs were shared between the three depth classes, and 55.1, 15.7 and 3.2% of the rarefied OTUs were unique to, respectively, the epi-, meso- and bathypelagic zones. Ciliates were by far the most abundant in terms of sequence reads. However, this observation is to be pondered by the fact that, in ciliate somatic macronuclei, rRNA genes are amplified several thousand times (e.g.

300 ~9000 copies in *Tetrahymena thermophila* (Ward et al., 1997)), such that their relative abundance in
term of cells is certainly much lower. Although diatoms (Bacillariophyta, Ochrophyta), several of them
considered endemic, are well known in Lake Baikal plankton (Moore et al., 2009; Zakharova et al., 2013;
Bashenkhaeva et al., 2015; Roberts et al., 2018; Mikhailov et al., 2019b), they represented only 6.1%
305 ochrophyte reads distributed in 64 OTUs. Optical microscopy on board showed that diatoms were
numerous, but their long frustules prevented most of them from being retained in the analyzed
plankton fraction. Chrysophytes represented the most abundant ochrophyte group, with 82.1% of
ochrophyte reads distributed in 698 OTUs. Members of the Holomycota were very diverse. Classical
fungi represented ca. 60% holomycotan sequences, most of them corresponding to chytrids, although
the Dicarya (Ascomycota, Basidiomycota) were relatively abundant too (Fig.S5). Most Dicyaria belonged
310 to typical terrestrial fungi entering the lake waters with river in-flow or from the surrounding land.
However, chytrids (flagellated fungi) are more likely to be truly planktonic organisms. Interestingly,
members of Rozellida (Cryptomycota) and Aphelida, were also relatively abundant, making up to almost
40% of the holomycotan sequences. Rozellids and aphelids, together with their microsporidian relatives
are parasites (Karpov et al., 2014; Bass et al., 2018). Although rozellids (cryptomycotes) are often
315 included within fungi, they are phagotrophic organisms, unlike fungi (which are osmotrophs), and they
branch more deeply than aphelids in the Holomycota tree (Torruella et al., 2018). Our data suggest that
the majority of actual fungal-like planktoners in Lake Baikal are parasites.

Overall, despite methodological differences, our identified plankton protist communities were
consistent with previous studies in surface waters or in a water column previously sampled in the
320 Southern basin, with ciliates, dinoflagellates and ochrophytes being highly represented (Annenkova et
al., 2020) (Yi et al., 2017).

Marine signature taxa

Although marine-freshwater transitions are thought to be rare (Mukherjee et al., 2019) and salinity, a
325 major driver of microbial community composition (Lozupone and Knight, 2007), high-throughput
environmental studies are revealing an increasing number of typically marine eukaryotic lineages in
freshwater systems. Among those are members of the parasitic perkinsids (Brate et al., 2010),
haptophytes (Simon et al., 2013), Bolidophyceae (Richards and Bass, 2005; Annenkova et al., 2020) and
several Marine Stramenopiles (MAST) clades (Massana et al., 2004; Massana et al., 2006), such as
330 MAST-2, MAST-12, MAST-3 and possibly MAST-6 (Simon et al., 2015a). Recently, diplomonads, a
cosmopolitan group of oceanic excavates particularly abundant and diverse in the deep ocean (Lara et
al., 2009; de Vargas et al., 2015) were identified in deep freshwater lakes (Yi et al., 2017; Mukherjee et
al., 2019). Likewise, Syndiniales, a clade of parasitic alveolates (often parasitizing their dinoflagellate

relatives) widely distributed in oceans (López-García et al., 2001; Guillou et al., 2008), were recently
335 identified in Baikal surface plankton (Annenkova et al., 2020). We identified members of all these
lineages in our large Lake Baikal plankton dataset, albeit mostly in low proportions (Fig.2A; Table S3).
Bolidophytes and, collectively, MAST clades were nonetheless relatively abundant in the lake. However,
MAST clades are not monophyletic and they exhibited different abundance patterns. Clades previously
detected in freshwater systems, MAST-2, MAST-6, MAST-12 and to a lesser extent MAST-3, were
340 relatively abundant. But MAST clades not previously observed in other freshwater systems, including
MAST-1, MAST-4, MAST-8 and MAST-20 occurred in very low proportions in a few samples. In addition
to the rare diplomonads, which were widely but sporadically present across Lake Baikal samples (Fig.
2A-B), we identified OTUs belonging to the emblematic Radiolaria, to our knowledge never before
identified in freshwater plankton. These OTUs were members of the Polycystinea (Fig.2C) and exhibited
345 extremely low frequencies.

The low abundance of some of these typically marine lineages partly explains the fact that they
failed to be detected in previous studies of freshwater systems, suggesting that these ecological
transitions have been so far underestimated (Paver et al., 2018). However, an additional explanation
might be found in the particular features of the Lake Baikal, including its considerable depth, marked
350 oligotrophy and even the presence of deep-venting (Müller et al., 2001; Sherstyankin et al., 2006;
UNDP-GEF, 2015), which make it qualify in all points but salinity as a freshwater sea.

Environmental drivers of protist community structure

To test whether planktonic protist communities were influenced by abiotic factors (clearly correlated
355 to sample spatial origin; Fig.1), we carried out permutational multivariate analysis of variance
(PERMANOVA) of Wisconsin-standardized Bray-Curtis distances between communities as a function of
sample spatial origin. PERMANOVA tests revealed significant differences in microbial eukaryotic
communities as a function of basin (latitudinal region), sampling site (coordinates) and depth within
sampling sites (Table 1). However, the most influential effects were those of the water column location,
360 23.7%, which combine latitudinal and vertical determinants, and depth within each single water
column, i.e. vertical variation alone (16.3%). The effect of the sampling basin was significant but small
(5.3%). To better visualize differences between communities, we carried out an NMDS analysis on the
global Bray-Curtis distance matrix. Points from most water columns did not show a marked
differentiation, as most water columns overlapped to some extent (Fig.S6). Likewise, samples from
365 different basins did not show a clear differentiation, although samples from the Southern basin tended
to segregate from the two other basins (Fig.3A). Samples from coastal versus open waters did not
segregate at all (Fig.3B). However, planktonic communities clearly segregated as a function of the water

column zonation, with epipelagic, mesopelagic and bathypelagic communities well separated in the NMDS plot (Fig.3C). NMDS analyses based on SWARM-defined OTUs yielded very similar results (Fig.S7).

370 These observations were statistically supported by ANOSIM tests, which showed significant and marked differences among communities according to depth, significant but weak differences according to basin origin, and no correlation at all between coastal and pelagic samples (Table S4).

These results suggest that depth is the major environmental factor structuring Lake Baikal protist communities. Depth is in turn a proxy for a variety of abiotic parameters, notably light, but also, despite
375 their limited variation, temperature, dissolved oxygen, conductivity and pH (Fig.1). These environmental variables and others, such as the nature of dissolved organic matter (TDS amount does not vary significantly; Fig.1B), are likely to influence prokaryotic communities as well (Kurilkina et al., 2016). Consequently, the nature of prey available for bacterivorous protists is possibly different. This may, in turn, select for protists with particular preying affinities, such that biotic interactions with other
380 planktonic members may be also important determinant factors of community structure and function.

Functional groups and biotic interactions

To look for potential ecological interactions between members of protist communities, we first explored the distribution of major functional classes with depth. We attributed protists to three major
385 categories based on knowledge about the lifestyle and ecological function of the corresponding phylogenetic lineages: autotrophs, free-living heterotrophs and parasites (Table S5). We acknowledge that these are very broad categories and that many photosynthetic organisms can be mixotrophs (Massana, 2011; Mitra et al., 2016). However, information about mixotrophy is still scarce and it is difficult to predict this ability from sequence data only. Therefore, our category 'autotrophs' included
390 also photosynthetic organisms that can additionally use heterotrophic feeding modes. Free-living heterotrophs include predatory protists but also osmotrophic organisms feeding on organic matter, such as fungi or some Stramenopiles. The relative abundance of the three functional categories in Lake Baikal significantly followed the same trend in the three water column zones, with autotrophs being less abundant than heterotrophs and parasites being in much lower proportion (Fig.4). Low proportions
395 of parasitic protists are consistent with affordable parasite loads for an ecosystem, as was previously observed (Simon et al., 2015b). Nonetheless, the relative amount of parasitic lineages diminished with depth, potentially suggesting that a relatively important proportion of protists identified in deep waters might be inactive. This is indeed likely the case for most photosynthetic organisms that were identified below the epipelagic region. Although the proportion of autotrophs diminished with depth, they still
400 made up to 30% of the total in bathypelagic waters, corresponding mostly to chrysophytes and dinoflagellates (respectively 43.7 and 45.1% of reads attributed to autotrophs). As mentioned, some of

these protists may be mixotrophic and prey on bacteria or other protists in the dark water column. However, the majority of photosynthetic lineages may simply be inactive, dormant or on their way to decay, serving as food for the heterotrophic component of microbial communities. The presence of relatively abundant photosynthetic protists in the Baikal dark water column and sediments is well documented (Zakharova et al., 2013; Yi et al., 2017), low temperatures possibly helping their preservation during sedimentation. Finally, free-living heterotrophs were the most abundant functional category throughout the water column. This might seem at odds with a pyramidal food-web structure whereby primary producers should be more abundant than consumers. However, several factors might explain this. First, the presence of ciliates likely introduces a positive bias in this functional category. Second, many autotrophs might be, on average, larger than heterotrophic protists and their biomass exceed that of consumers, although some studies suggest that relative OTU proportions reflect better biomass than cell counts (Egge et al., 2013; Bernstein et al., 2017). Finally, many heterotrophic protists might depend on bacteria or on larger organisms (e.g. fungi degrading decaying plant material) that have been excluded from the studied size range.

To further explore biotic interactions, and given that protist community differences were essentially seen throughout the water column, we reconstructed co-occurrence networks of OTUs found in epipelagic, mesopelagic and bathypelagic zones. To build the networks, we retained OTUs present in more than 20% of samples at relative abundances higher than 0.01% (Table S6). The structure of the three networks was markedly different (Fig.5). The epipelagic network was denser, having more interconnected OTUs, more positive interactions and several hub-type OTUs that interact with many OTUs. Mean node degrees were also higher in the epipelagic network (Table S7). Meso- and bathypelagic networks had less connected nodes and most correlations were negative. Although bathypelagic waters exhibited the least connected nodes, both the connectance and the clustering coefficient of the network were the highest. However, the number of epipelagic samples was higher than those for the two other depth zones. Therefore, to test whether differences in the number of samples between the three depth classes could explain the observed differences in connectivity, we performed a bootstrap analysis, generating networks based on the same number of samples taken randomly for each depth. This showed that network connectivity properties were not significantly different for epipelagic and, respectively, mesopelagic or bathypelagic samples (Fig. S8.A-B). However, when we merged samples from the aphotic zone (meso- and bathypelagic samples), the bootstrap analysis revealed that network connectivity properties were significantly higher for the surface samples than for samples from the dark water column (Fig. S8.C; Table S7). A block-model representation of the three networks indicated the occurrence of pairs of OTU sets sharing similar properties that were highly interconnected with each other and only loosely to other sets (Fig.S9). The epipelagic block-model

representation translates well the higher interconnection of community members, with four blocks of OTUs displaying an important number of connections with at least another block. In terms of interacting members, positive interactions (co-occurrence) involved one dinoflagellate, one chryptophyte and several ochrophytes (all potentially photosynthetic) in interaction with ciliate OTUs in epipelagic waters (Fig. 5A). This might suggest specific predation of ciliates on particular algal species. In deeper waters, significant co-occurrences were rare. Only one positive interaction was observed in mesopelagic waters (ciliate-fungus) and, similarly, in bathypelagic waters (rozellid-ochrophyte). The latter might suggest a specific parasitic interaction. Negative correlations (co-exclusions) are more difficult to interpret and might reflect systematic preference for different conditions between sites.

Collectively, our network data suggest that epipelagic communities are more active and have more positive and negative interactions, whereas in the deep, dark water column, communities are more stable with a restricted connected core.

Concluding remarks

Lake Baikal in Southern Siberia is a unique freshwater system by its volume, maximum depth (1,642 m) and topographical features that include rifting associated with hydrothermalism. With its highly oligotrophic waters, it amounts to an inner freshwater sea in all points but an extremely low salinity. Freshwater ecosystems are particularly threatened by climate change and, being located in Southern Siberia, one of the most rapidly changing zones, Lake Baikal is being severely impacted (Mackay et al., 2017). Yet, despite the importance of the lake and its uniqueness, its microbial planktonic communities have been only partially studied and we lack reference comprehensive comparative community data to assess ongoing and future change and infer how it may affect microbial functions and the ecology of the lake. In this study, we have analyzed the composition of microbial eukaryotic communities in plankton collected from different water columns along a transect of ~600 km across the three lake basins, from surface (5 m) to high depth (1,400 m) and from littoral to open waters. Our study shows widely diverse communities covering all eukaryotic supergroups, with ciliates, dinoflagellates, chrysophytes and flagellated fungi plus related lineages (rozellids, aphelids) being the most relatively abundant, together with cryptophytes, haptophytes, katablepharids, telonemids and several MAST lineages. Interestingly, confirming previous observations in Lake Baikal, we observed members of typically marine lineages, including bolidophytes, syndineans, diplonemids and, for the first time, radiolarians. These observations suggest that the salinity barrier is relatively easy to cross and that the 'marine' determinants might be more related with the oligotrophic nature of the system and the occurrence of a deep water column than with salinity itself. Despite the relatively homogeneous values of several physicochemical parameters, planktonic protist communities were highly and significantly

470 stratified in Lake Baikal, suggesting that depth, as a proxy for light but also temperature, pH, oxygen
and conductivity, is a major determinant of community structure. By contrast, the effect of latitude
(basins) was minor, if not negligible. Consistent with vertical stratification, the relative proportion of
autotrophs, free-living heterotrophs and parasites is altered with depth, where photosynthetic lineages
are still present but, like parasites, in lower proportions. Biotic factors are also important in structuring
475 Lake Baikal communities. Co-occurrence network analyses showed highly interconnected communities
in surface waters, with positive and negative interactions. By contrast, communities from the dark
water column exhibit much less connected OTUs, although they are strongly, and mostly negatively,
correlated. This might be suggestive of much more diluted and potentially inactive populations, but
with a conserved core of highly interconnected OTUs. Our results pave the way for future comparative
480 analyses of protist communities through time, notably in the context of rapid climate change that is
affecting Siberia and Lake Baikal.

Acknowledgments

485 We thank the crew of the R/V G. Titov for their professionalism and efficiency onboard, the director of
the Limnological Institute at Irkusk for logistical assistance and Philippe Deschamps for technical
bioinformatic support. This research was funded by the European Research Council Grants ProtistWorld
(322669, PL-G) and PlastEvol (787904, DM) as well as the Russian State grant 0345-2016-0009 (NVA).

490 Author contributions

PLG, DM and NVA designed the work and organized the limnological cruise. PLG, PB, AILA, LG, GR and
NVA collected and processed water samples during the cruise. PB purified DNA and carried out PCR
reactions for metabarcoding analysis. GR carried out the initial bioinformatic analysis of amplicon
sequences. GD carried out metabarcoding, statistical and network analyses, with help from LJ. PLG
495 wrote the manuscript with input from co-authors. All authors read, critically commented and approved
the final manuscript.

Conflict of interest

The authors declare that they have no conflicts of interest.

500

References

Adams, J.K., Martins, C.C., Rose, N.L., Shchetnikov, A.A., and Mackay, A.W. (2018) Lake sediment
records of persistent organic pollutants and polycyclic aromatic hydrocarbons in southern Siberia
mirror the changing fortunes of the Russian economy over the past 70 years. *Environ Pollut* **242**:
505 528-538.

- Annenkova, N.V., Lavrov, D.V., and Belikov, S.I. (2011) Dinoflagellates associated with freshwater sponges from the ancient lake baikal. *Protist* **162**: 222-236.
- Annenkova, N.V., Giner, C.R., and Logares, R. (2020) Tracing the origin of planktonic protists in an ancient lake. *Microorganisms* **8**.
- 510 Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T., and Stamatakis, A. (2019) EPA-ng: Massively parallel evolutionary placement of genetic sequences. *Syst Biol* **68**: 365-369.
- Barrat, A., Barthélemy, M., and Vespignani, A. (2008) *Dynamical Processes on Complex Networks*. Cambridge: Cambridge University Press.
- Bashenkhaeva, M.V., Zakharova, Y.R., Petrova, D.P., Khanaev, I.V., Galachyants, Y.P., and Likhoshway, Y.V. (2015) Sub-ice microalgal and bacterial communities in freshwater Lake Baikal, Russia. *Microb Ecol* **70**: 751-765.
- 515 Bass, D., Czech, L., Williams, B.A.P., Berney, C., Dunthorn, M., Mahe, F. et al. (2018) Clarifying the relationships between Microsporidia and Cryptomycota. *J Eukaryot Microbiol* **65**: 773-782.
- Bel'kova, N.L., Parfenova, V.V., Kostopnova, T., Denisova, L., and Zaichikov, E.F. (2003) [Microbial biodiversity in the Lake Baikal water]. *Mikrobiologiya* **72**: 239-249.
- 520 Belikov, S., Belkova, N., Butina, T., Chernogor, L., Martynova-Van Kley, A., Nalian, A. et al. (2019) Diversity and shifts of the bacterial community associated with Baikal sponge mass mortalities. *PLoS One* **14**: e0213926.
- Bernstein, H.C., Brislawn, C.J., Dana, K., Flores-Wentz, T., Cory, A.B., Fansler, S.J. et al. (2017) Primary and heterotrophic productivity relate to multikingdom diversity in a hypersaline mat. *FEMS Microbiol Ecol* **93**.
- 525 Bower, S.M., Carnegie, R.B., Goh, B., Jones, S.R.M., Lowe, G.J., and Mak, M.W.S. (2004) Preferential PCR amplification of parasitic protistan small subunit rDNA from metazoan tissues. *J Euk Microbiol* **51**: 325-332.
- 530 Brate, J., Logares, R., Berney, C., Ree, D.K., Klaveness, D., Jakobsen, K.S., and Shalchian-Tabrizi, K. (2010) Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *ISME J* **4**: 1144-1153.
- Burki, F., Roger, A.J., Brown, M.W., and Simpson, A.G.B. (2020) The New Tree of Eukaryotes. *Trends in Ecology & Evolution* **35**: 43-55.
- 535 Butina, T.V., Bukin, Y.S., Krasnopeev, A.S., Belykh, O.I., Tupikin, A.E., Kabilov, M.R. et al. (2019) Estimate of the diversity of viral and bacterial assemblage in the coastal water of Lake Baikal. *FEMS Microbiol Lett* **366**.
- Cabello-Yeves, P.J., Zemskaia, T.I., Rosselli, R., Coutinho, F.H., Zakharenko, A.S., Blinov, V.V., and Rodriguez-Valera, F. (2018) Genomes of Novel Microbial Lineages Assembled from the Sub-Ice Waters of Lake Baikal. *Appl Environ Microbiol* **84**.
- 540 Cabello-Yeves, P.J., Zemskaia, T.I., Zakharenko, A.S., Sakirko, M.V., Ivanov, V.G., Ghai, R., and Rodriguez-Valera, F. (2020) Microbiome of the deep Lake Baikal, a unique oxic bathypelagic habitat. *Limnol Oceanogr* **n/a**.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**: 581-583.
- 545 Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-1973.
- Caron, D.A., Countway, P.D., Savai, P., Gast, R.J., Schnetzer, A., Moorthi, S.D. et al. (2009) Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Appl Environ Microbiol* **75**: 5797-5808.
- 550 Cavicchioli, R., Ripple, W.J., Timmis, K.N., Azam, F., Bakken, L.R., Baylis, M. et al. (2019) Scientists' warning to humanity: microorganisms and climate change. *Nat Rev Microbiol* **17**: 569-586.
- Chiquet, J., Mariadassou, M., and Robin, S. (2018) Variational inference for probabilistic Poisson PCA. *Ann Appl Stat* **12**: 2674-2698.
- 555 Csardi, G., and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal, Complex Systems* **1695**: 1-9.

- De Batist, M., Klerkx, J., Van Rensbergen, P., Vanneste, M., Poort, J., Golmshtok, A.Y. et al. (2002) Active hydrate destabilization in Lake Baikal, Siberia? *Terra Nova* **14**: 436-442.
- 560 de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R. et al. (2015) Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**: 1261605.
- Delmas, E., Besson, M., Brice, M.-H., Burkle, L.A., Dalla Riva, G.V., Fortin, M.-J. et al. (2019) Analysing ecological networks of species interactions. *Biological Reviews* **94**: 16-36.
- Dray, S., and Dufoour, A.B. (2007) The ade4 Package : Implementing the duality diagram for ecologists. *Journal of Statistical Software* **22-4**.
- 565 Egge, E., Bittner, L., Andersen, T., Audic, S., de Vargas, C., and Edvardsen, B. (2013) 454 pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. *PLoS One* **8**: e74371.
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150-3152.
- 570 Glöckner, F.O., Zaichikov, E., Belkova, N., Denissova, L., Pernthaler, J., Pernthaler, A., and Amann, R. (2000) Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria. *Appl Environ Microbiol* **66**: 5053-5065.
- 575 Granin, N.G., Aslamov, I.A., Kozlov, V.V., Makarov, M.M., Kirillin, G., McGinnis, D.F. et al. (2019) Methane hydrate emergence from Lake Baikal: direct observations, modelling, and hydrate footprints in seasonal ice cover. *Sci Rep* **9**: 19361.
- 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.
- 580 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-D604.
- Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696-704.
- 585 Karpov, S.A., Mamkaeva, M.A., Aleoshin, V.V., Nassonova, E., Lilje, O., and Gleason, F.H. (2014) Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Front Microbiol* **5**: 112.
- Kashiwaya, K., Ochiai, S., Sakai, H., and Kawai, T. (2001) Orbit-related long-term climate cycles revealed in a 12-Myr continental record from Lake Baikal. *Nature* **410**: 71-74.
- 590 Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772-780.
- Kurilkina, M.I., Zakharova, Y.R., Galachyants, Y.P., Petrova, D.P., Bukin, Y.S., Domysheva, V.M. et al. (2016) Bacterial community composition in the water column of the deepest freshwater Lake Baikal as determined by next-generation sequencing. *FEMS Microbiol Ecol* **92**.
- 595 Lara, E., Moreira, D., Vereshchaka, A., and Lopez-Garcia, P. (2009) Pan-oceanic distribution of new highly diverse clades of deep-sea diplomonads. *Environ Microbiol* **11**: 47-55.
- Lê, S., Josse, J., and Husson, F. (2008) FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software* **25**: 1-18.
- Leger, J.B. (2016) Blockmodels: A R-package for estimating in Latent Block Model and Stochastic Block Model, with various probability functions, with or without covariates. <http://arxiv.org/abs/160207587>.
- 600 López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C., and Moreira, D. (2001) Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**: 603-607.
- Lozupone, C.A., and Knight, R. (2007) Global patterns in bacterial diversity. *Proc Natl Acad Sci U S A* **104**: 605 11436-11440.

- Mackay, A.W., Seddon, A.W., Leng, M.J., Heumann, G., Morley, D.W., Piotrowska, N. et al. (2017) Holocene carbon dynamics at the forest-steppe ecotone of southern Siberia. *Glob Chang Biol* **23**: 1942-1960.
- 610 Magoc, T., and Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**: 2957-2963.
- Mahe, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. (2015) Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* **3**: e1420.
- 615 Maksimov, V.V., Shchetinina, E.V., Kraikovskaia, O.V., Maksimov, V.N., and Maksimova, E.A. (2002) [The classification and the monitoring of the state of mouth riverine and lacustrine ecosystems in lake Baikal based on the composition of local microbiocenoses and their activity]. *Mikrobiologiya* **71**: 690-696.
- Maksimova, E.A., and Maksimov, V.N. (1972) [Vertical distribution of microbial plankton in the Southern part of Lake Baikal in 1969]. *Mikrobiologiya* **41**: 896-902.
- 620 Markovic, D., Carrizo, S.F., Kärcher, O., Walz, A., and David, J.N.W. (2017) Vulnerability of European freshwater catchments to climate change. *Glob Chang Biol* **23**: 3567-3580.
- Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJournal* **17**: 10-12.
- Massana, R. (2011) Eukaryotic picoplankton in surface oceans. *Annu Rev Microbiol* **65**: 91-110.
- 625 Massana, R., Terrado, R., Forn, I., Lovejoy, C., and Pedros-Alio, C. (2006) Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. *Environ Microbiol* **8**: 1515-1522.
- Massana, R., Castresana, J., Balague, V., Guillou, L., Romari, K., Groisillier, A. et al. (2004) Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl Environ Microbiol* **70**: 3528-3534.
- Mats, V.D., and Perepelova, T.I. (2011) A new perspective on evolution of the Baikal Rift. *Geoscience Frontiers* **2**: 349-365.
- 630 Mikhailov, I.S., Zakharova, Y.R., Bukin, Y.S., Galachyants, Y.P., Petrova, D.P., Sakirko, M.V., and Likhoshway, Y.V. (2019a) Co-occurrence networks among bacteria and microbial eukaryotes of Lake Baikal during a spring phytoplankton bloom. *Microb Ecol* **77**: 96-109.
- Mikhailov, I.S., Zakharova, Y.R., Galachyants, Y.P., Usoltseva, M.V., Petrova, D.P., Sakirko, M.V. et al. (2015) Similarity of structure of taxonomic bacterial communities in the photic layer of Lake Baikal's three basins differing in spring phytoplankton composition and abundance. *Dokl Biochem Biophys* **465**: 413-419.
- 635 Mikhailov, I.S., Bukin, Y.S., Zakharova, Y.R., Usoltseva, M.V., Galachyants, Y.P., Sakirko, M.V. et al. (2019b) Co-occurrence patterns between phytoplankton and bacterioplankton across the pelagic zone of Lake Baikal during spring. *J Microbiol* **57**: 252-262.
- 640 Mitra, A., Flynn, K.J., Tillmann, U., Raven, J.A., Caron, D., Stoecker, D.K. et al. (2016) Defining planktonic protist functional groups on mechanisms for energy and nutrient acquisition: incorporation of diverse mixotrophic strategies. *Protist* **167**: 106-120.
- Moore, M.V., Hampton, S.E., Izmet'seva, L.R., Silow, E.A., Peshkova, E.V., and Pavlov, B.K. (2009) Climate change and the world's "Sacred Sea"—Lake Baikal, Siberia. *BioScience* **59**: 405-417.
- 645 Mukherjee, I., Hodoki, Y., Okazaki, Y., Fujinaga, S., Ohbayashi, K., and Nakano, S.I. (2019) Widespread dominance of kinetoplastids and unexpected presence of diplomonads in deep freshwater lakes. *Front Microbiol* **10**: 2375.
- Müller, J., Oberhänsli, H., Melles, M., Schwab, M., Rachold, V., and Hubberten, H.W. (2001) Late Pliocene sedimentation in Lake Baikal: implications for climatic and tectonic change in SE Siberia. *Palaeogeogr Palaeoclimatol Palaeoecol* **174**: 305-326.
- 650 Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268-274.
- Okamoto, N., Chantangsi, C., Horak, A., Leander, B.S., and Keeling, P.J. (2009) Molecular phylogeny and description of the novel katablepharid *Roombia truncata* gen. et sp. nov., and establishment of the *Hacrobia* taxon nov. *PLoS ONE* **4**: e7080.
- 655

- Oksanen, J., Blanchet, G., Kindt, R., Legendre, P., O'Hara, R.B., Simpson, G.L. et al. (2011) Vegan: Community Ecology Package. R package version 1.17-9. In. <http://CRAN.R-project.org/package=vegan> (ed): <http://CRAN.R-project.org/package=vegan>.
- 660 Paver, S.F., Muratore, D., Newton, R.J., and Coleman, M.L. (2018) Reevaluating the salty divide: phylogenetic specificity of transitions between marine and freshwater systems. *mSystems* **3**.
- Pielou, E.C. (1966) Species-diversity and pattern-diversity in the study of ecological succession. *J Theor Biol* **10**: 370-383.
- Potapov, S.A., Tikhonova, I.V., Krasnopeev, A.Y., Kabilov, M.R., Tupikin, A.E., Chebunina, N.S. et al. (2019) Metagenomic analysis of virioplankton from the pelagic zone of Lake Baikal. *Viruses* **11**.
- 665 Prokopenko, A.A., Karabanov, E.B., and Williams, D.F. (2002) Age of long sediment cores from Lake Baikal. *Nature* **415**: 976.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590-596.
- 670 R Development Core Team (2017) R: A language and environment for statistical computing. In. <http://www.r-project.org> (ed). Vienna, Austria: R Foundation for Statistical Computing.
- Richards, T.A., and Bass, D. (2005) Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. *Curr Opin Microbiol* **8**: 240-252.
- Roberts, S., Adams, J.K., Mackay, A.W., Swann, G.E.A., McGowan, S., Rose, N.L. et al. (2020) Mercury loading within the Selenga River basin and Lake Baikal, Siberia. *Environ Pollut* **259**: 113814.
- 675 Roberts, S.L., Swann, G.E.A., McGowan, S., Panizzo, V.N., Vologina, E.G., Sturm, M., and Mackay, A.W. (2018) Diatom evidence of 20th century ecosystem change in Lake Baikal, Siberia. *PLoS One* **13**: e0208765.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahe, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**: e2584.
- 680 Rolls, R.J., Heino, J., Ryder, D.S., Chessman, B.C., Gowns, I.O., Thompson, R.M., and Gido, K.B. (2018) Scaling biodiversity responses to hydrological regimes. *Biol Rev Camb Philos Soc* **93**: 971-995.
- Schmid, M., Budnev, N.M., Granin, N.G., Sturm, M., Schurter, M., and Wüest, A. (2008) Lake Baikal deepwater renewal mystery solved. *Geophys Res Lett* **35**: L09605-L09605.
- 685 Schuur, E.A., McGuire, A.D., Schädel, C., Grosse, G., Harden, J.W., Hayes, D.J. et al. (2015) Climate change and the permafrost carbon feedback. *Nature* **520**: 171-179.
- Sherstyankin, P.P., Alekseev, S.P., Abramov, A.M., Stavrov, K.G., De Batist, M., Hus, R. et al. (2006) Computer-based bathymetric map of Lake Baikal. *Dokl Earth Sci* **408**: 564-569.
- Shimaraev, M.N., and Domysheva, V.M. (2012) Trends in hydrological and hydrochemical processes in Lake Baikal under conditions of modern climate change. In. Chichester, UK: John Wiley & Sons, Ltd, pp. 43-66.
- 690 Simon, M., Lopez-Garcia, P., Moreira, D., and Jardillier, L. (2013) New haptophyte lineages and multiple independent colonizations of freshwater ecosystems. *Environ Microbiol Rep* **5**: 322-332.
- Simon, M., Jardillier, L., Deschamps, P., Moreira, D., Restoux, G., Bertolino, P., and Lopez-Garcia, P. (2015a) Complex communities of small protists and unexpected occurrence of typical marine lineages in shallow freshwater systems. *Environ Microbiol* **17**: 3610-3627.
- 695 Simon, M., Lopez-Garcia, P., Deschamps, P., Moreira, D., Restoux, G., Bertolino, P., and Jardillier, L. (2015b) Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. *ISME J* **9**: 1941-1953.
- 700 Torruella, G., Grau-Bove, X., Moreira, D., Karpov, S.A., Burns, J.A., Sebe-Pedros, A. et al. (2018) Global transcriptome analysis of the aphelid *Paraphelidium tribonemae* supports the phagotrophic origin of fungi. *Commun Biol* **1**: 231.
- Troitskaya, E., Blinov, V., Ivanov, V., Zhdanov, A., Gnatovsky, R., Sutyryna, E., and Shimaraev, M. (2015) Cyclonic circulation and upwelling in Lake Baikal. *Aquatic Sciences* **77**: 171-182.
- 705 UNDP-GEF (2015) The ecological atlas of the Baikal basin. In. <http://baikal.iwlearn.org/en>: United Nations Office for Project Sercives (UNOPS), p. 145.

- Ward, J.G., Blomberg, P., Hoffman, N., and Yao, M.C. (1997) The intranuclear organization of normal, hemizygous and excision-deficient rRNA genes during developmental amplification in *Tetrahymena thermophila*. *Chromosoma* **106**: 233-242.
- 710 Wilburn, P., Shchapov, K., Theriot, E.C., and Litchman, E. (2020) Environmental drivers define contrasting microbial habitats, diversity, and community structure in Lake Baikal, Siberia. *bioRxiv*: 605899.
- Woodward, G., Perkins, D.M., and Brown, L.E. (2010) Climate change and freshwater ecosystems: impacts across multiple levels of organization. *Philos Trans R Soc Lond B Biol Sci* **365**: 2093-2106.
- 715 Yi, Z., Berney, C., Hartikainen, H., Mahamdallie, S., Gardner, M., Boenigk, J. et al. (2017) High-throughput sequencing of microbial eukaryotes in Lake Baikal reveals ecologically differentiated communities and novel evolutionary radiations. *FEMS Microbiol Ecol* **93**: 10.
- Zakharenko, A.S., Galachyants, Y.P., Morozov, I.V., Shubenkova, O.V., Morozov, A.A., Ivanov, V.G. et al. (2019) Bacterial communities in areas of oil and methane seeps in pelagic of Lake Baikal. *Microb Ecol*
- 720 **78**: 269-285.
- Zakharova, Y.R., Galachyants, Y.P., Kurilkina, M.I., Likhoshvay, A.V., Petrova, D.P., Shishlyannikov, S.M. et al. (2013) The structure of microbial community and degradation of diatoms in the deep near-bottom layer of Lake Baikal. *PLoS One* **8**: e59977.

725

730

Figure Legends

735 **Fig. 1.** Sampling sites and overall planktonic protist community composition in Lake Baikal. **A**, map of Lake Baikal showing the different sampling sites across the three major lake basins (indicated by colors). **B**, Principal component analysis (PCA) of samples according to their associated physicochemical parameters. TDS, total dissolved solids; DO, dissolved oxygen; ORP, oxidation-reduction potential. Blue tones indicate the sampling depth in the water column, as indicated. **C**, Relative abundance of different
740 high-rank eukaryotic taxa in Baikal plankton based on read counts for the defined OTUs. The asterisk indicates samples retrieved from Cell-Traps (Methods). Color codes for sample basin and depth origin as well as for the different taxa are indicated.

Fig. 2. Marine signature taxa detected in Lake Baikal plankton. **A**, heat map showing the relative
745 abundance of different typically marine taxa across Baikal plankton samples. The frequency of the different phylogenetic groups is indicated by different shades of blue. **B**, Maximum Likelihood (ML) phylogenetic tree of OTUs belonging to diplomonads and a related group of euglenozoan excavates (594 unambiguously aligned positions). **C**, ML tree of radiolarian OTUs (534 unambiguously aligned positions).

750 **Fig. 3.** Non-metric multidimensional scaling (NMDS) analysis of Lake Baikal plankton samples as a function of protist community similarities. The NMDS plot was constructed with Wisconsin-standardized Bray-Curtis dissimilarities between all samples. **A**, plankton samples highlighted by basin origin. **B**, plankton samples from coastal, shallow sites versus open water sites. **C**, samples grouped according to their depth origin in the water column; epipelagic (<200 m), mesopelagic (200-1,000 m),
755 bathypelagic (>1,000 m).

Fig. 4. Box plots showing the distribution of relative abundances of major functional categories of protists in Lake Baikal plankton. The three plots show the relative abundance of sequences affiliated to
760 autotrophic, heterotrophic and parasitic protists for each sampling depth class. The thickest line inside each box represents the median on the distribution, bottom and top borders of boxes correspond to the first and the third quartiles and whiskers extend to minimal and maximal distances. Significant differences between distributions are indicated with stars (p -values <0.05, <0.005 and <0.0005 are respectively indicated by one, two and three stars). For the assignation of taxa to functional categories, see Supplementary Table 5.
765

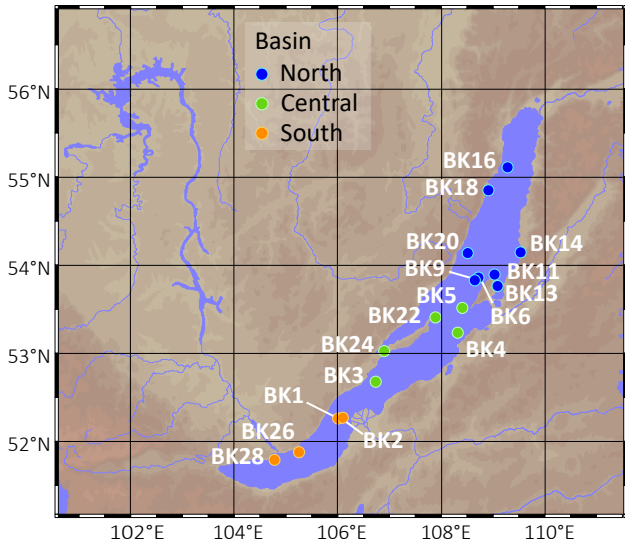
Fig. 5. Co-occurrence networks of planktonic protists in the Lake Baikal water column. A, network obtained from epipelagic (<200 m) samples across the lake. B, network obtained from mesopelagic (200-1,000 m) samples. C, network obtained from bathypelagic (>1,000 m samples). Networks were built on OTUs present in more than 20% samples and having a relative abundance higher than 0.01%.

770 OTUs are represented by nodes and direct covariations between them, by edges. Nodes and taxa labeled with an asterisk correspond to putative parasites.

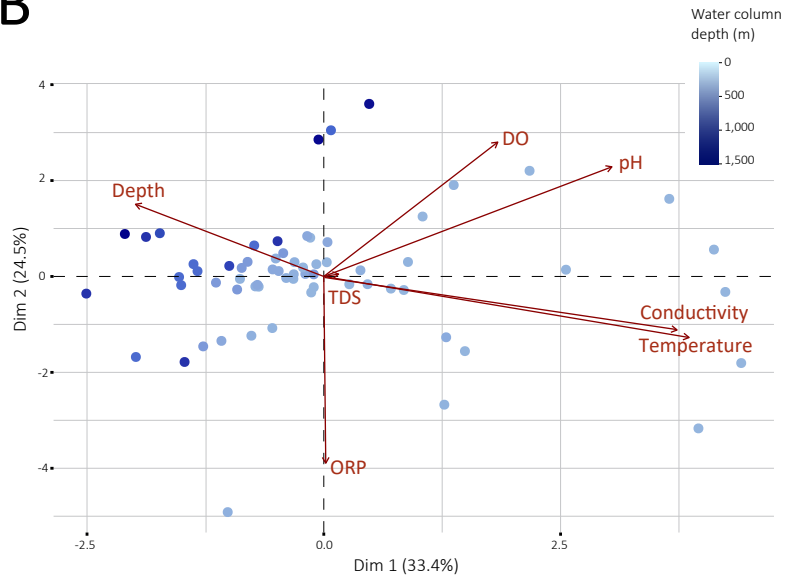
775 **Table 1.** Permutational multivariate analyse of variance (PERMANOVA) of Lake Baikal planktonic protist communities across basins, sampling site and depth. PERMANOVA was calculated using Wisconsin standardization on rarefied OTUs belonging to the 65 studied plankton samples. Df, degrees of freedom.

Effect	Df	F.Model	R ²	P value
Region	2	1.8369	5.30%	***
Sampling site	15	1.1857	23.70%	***
Depth Sampling site	9	1.2861	16.30%	***

A



B



C



Figure 1. G. David et al.

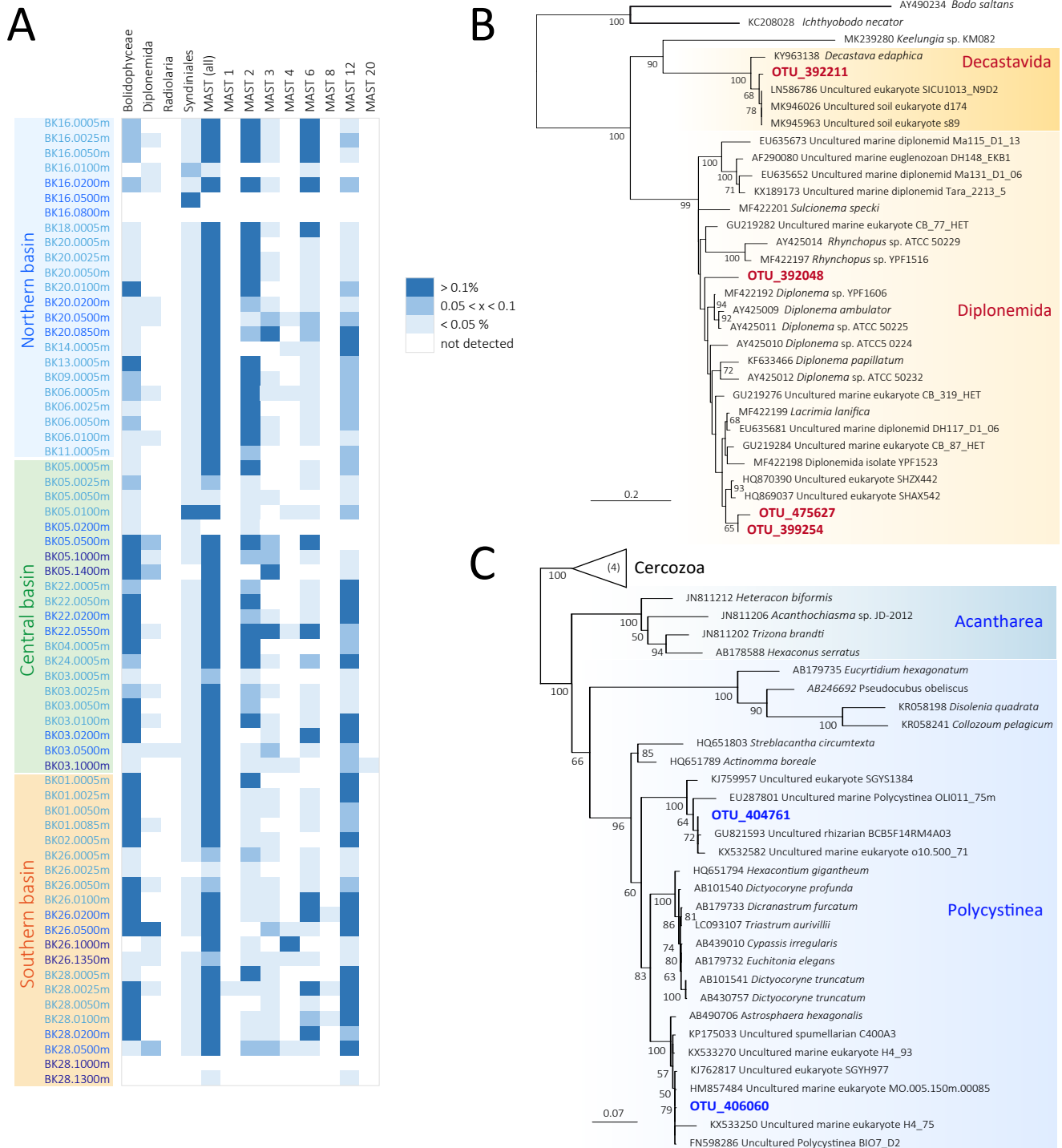


Figure 2. G. David et al.

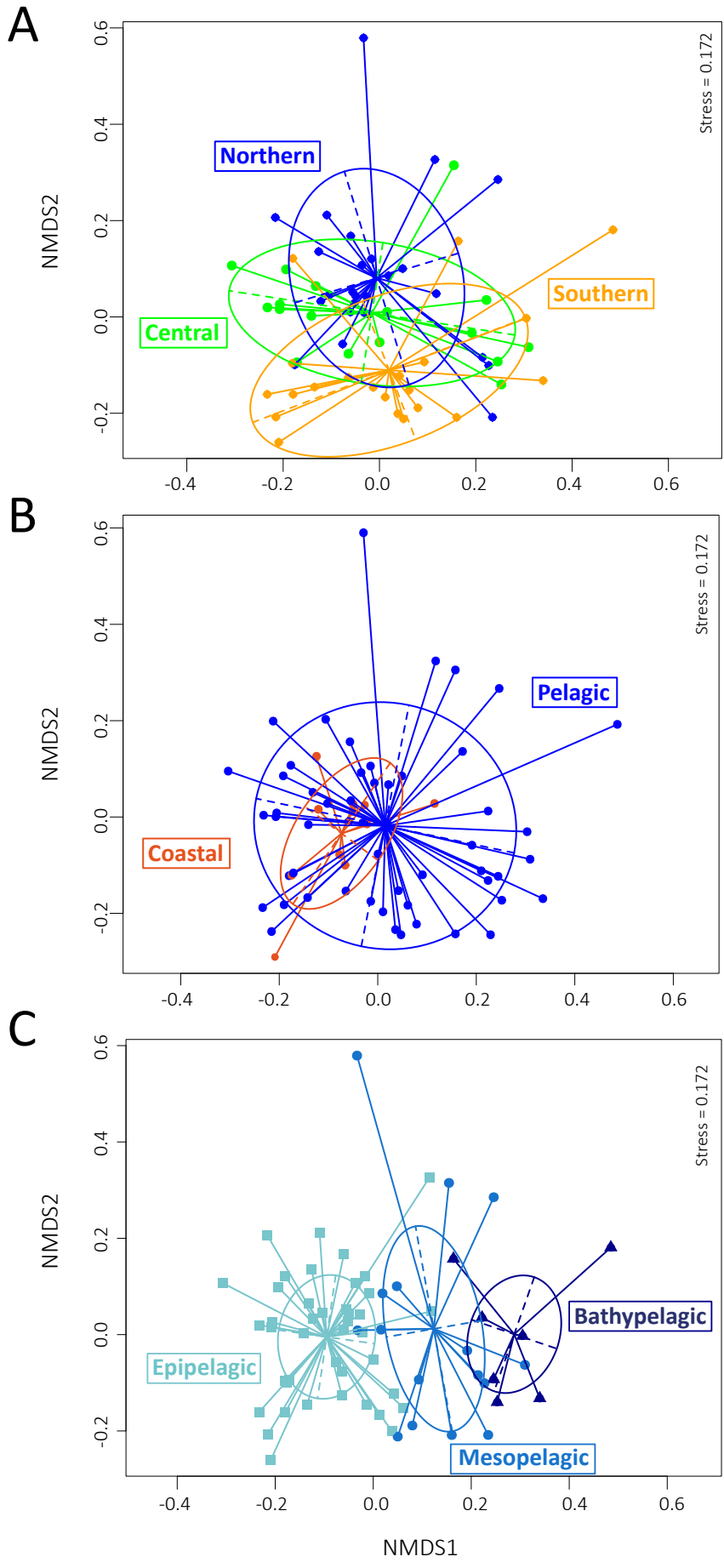


Figure 3. G. David et al.

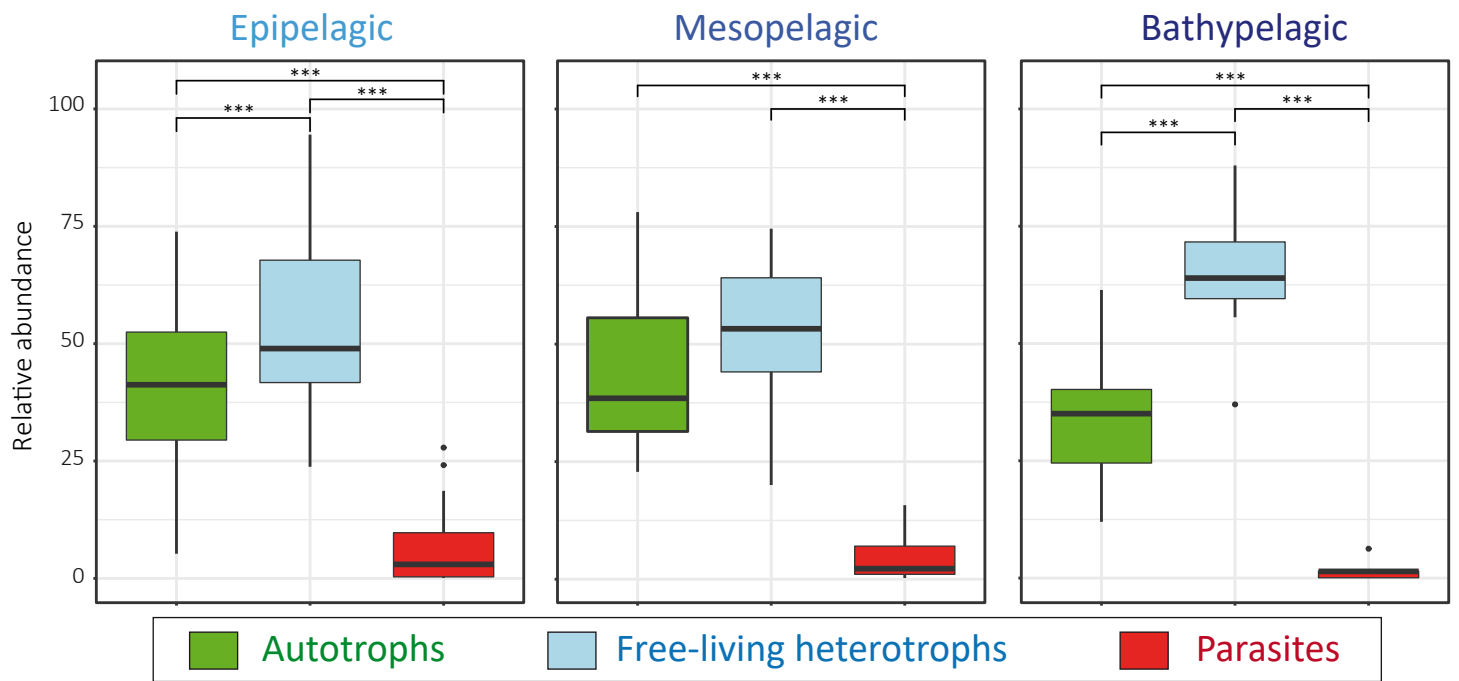
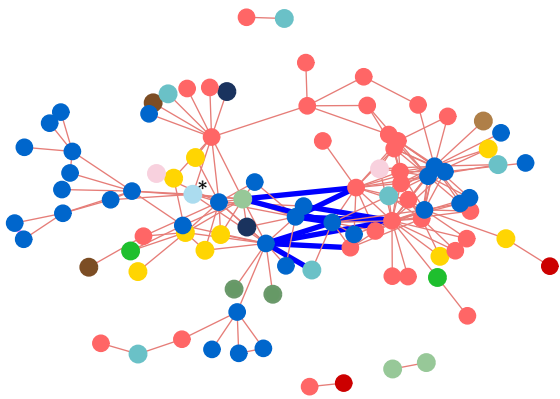
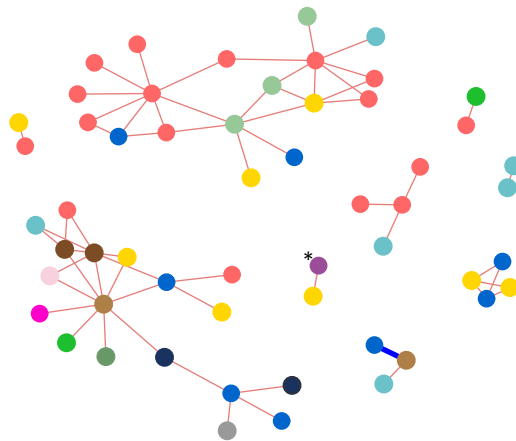


Figure 4. G. David et al.

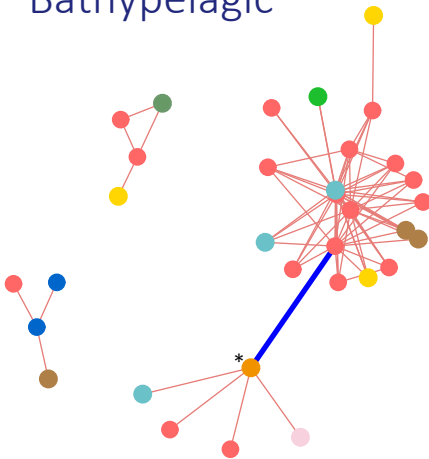
A Epipelagic



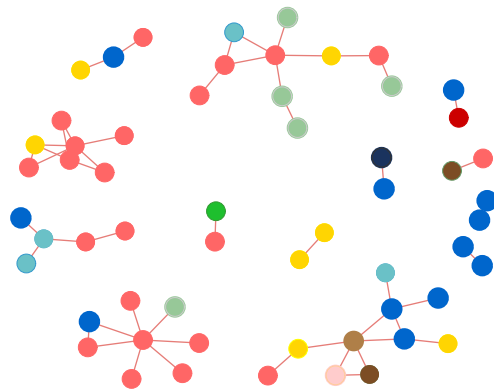
B Mesopelagic



C Bathypelagic



D Meso + bathypelagic



— Positive edge
— Negative edge

- Ciliophora
- Dinophyta
- Perkinsia*
- Chlorophyta
- Diplonemida
- Cryptophyta
- Haptophyta
- Katablepharida
- Telonemia
- Choanoflagellida
- Fungi
- Rozellida/Cryptomycota*
- Aphelida*
- Cercozoa
- MAST
- Ochrophyta
- Other Stramenopiles

Figure 5. G. David et al.

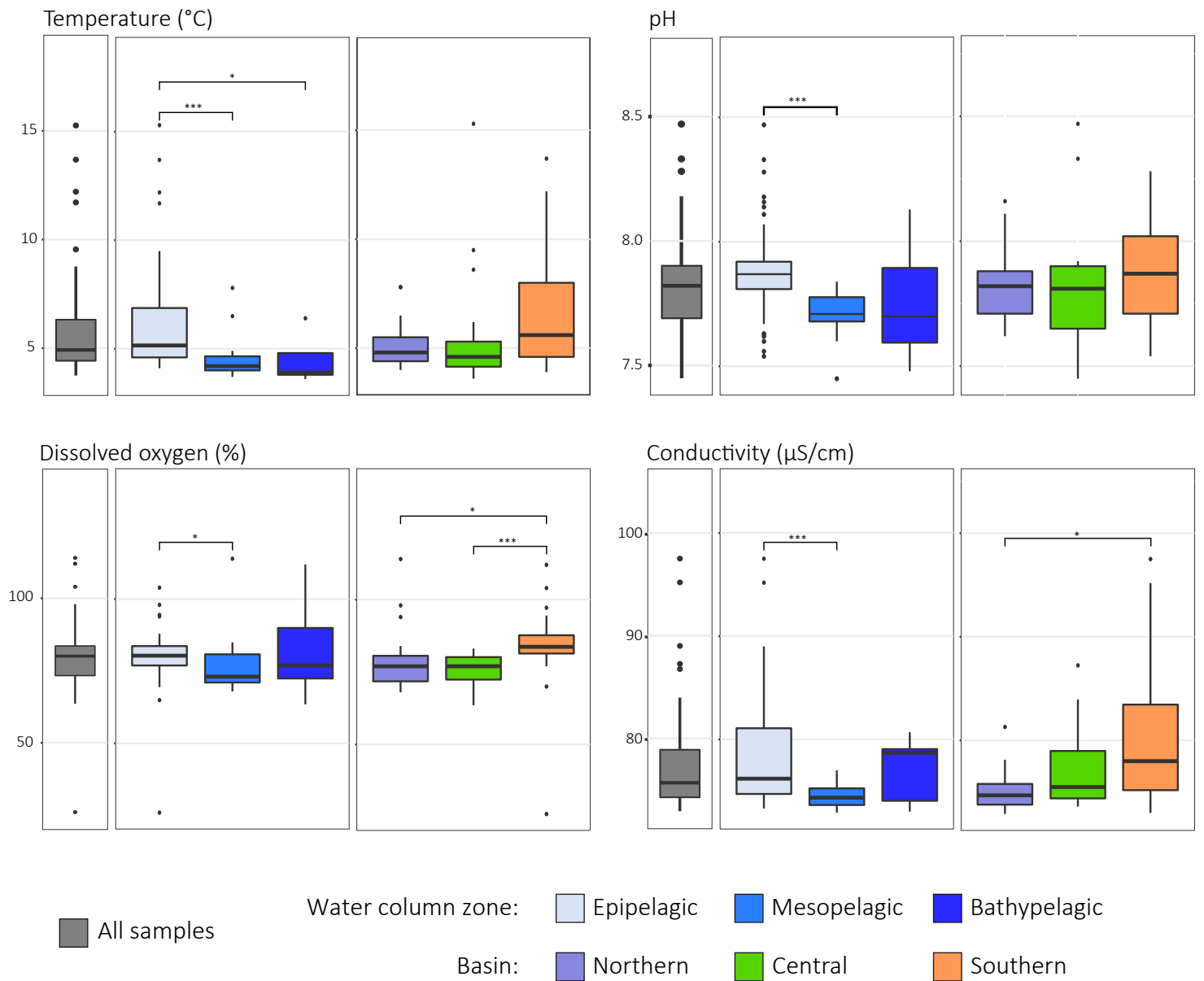


Figure S1. Box plots of physico-chemical parameters measured in Lake Baikal water samples. Average values and variation are shown for all the Baikal samples and as a function of basin and depth. Significant differences between pairs of samples are indicated (p-value ≤ 0.05 *; p-value < 0.001 ***).

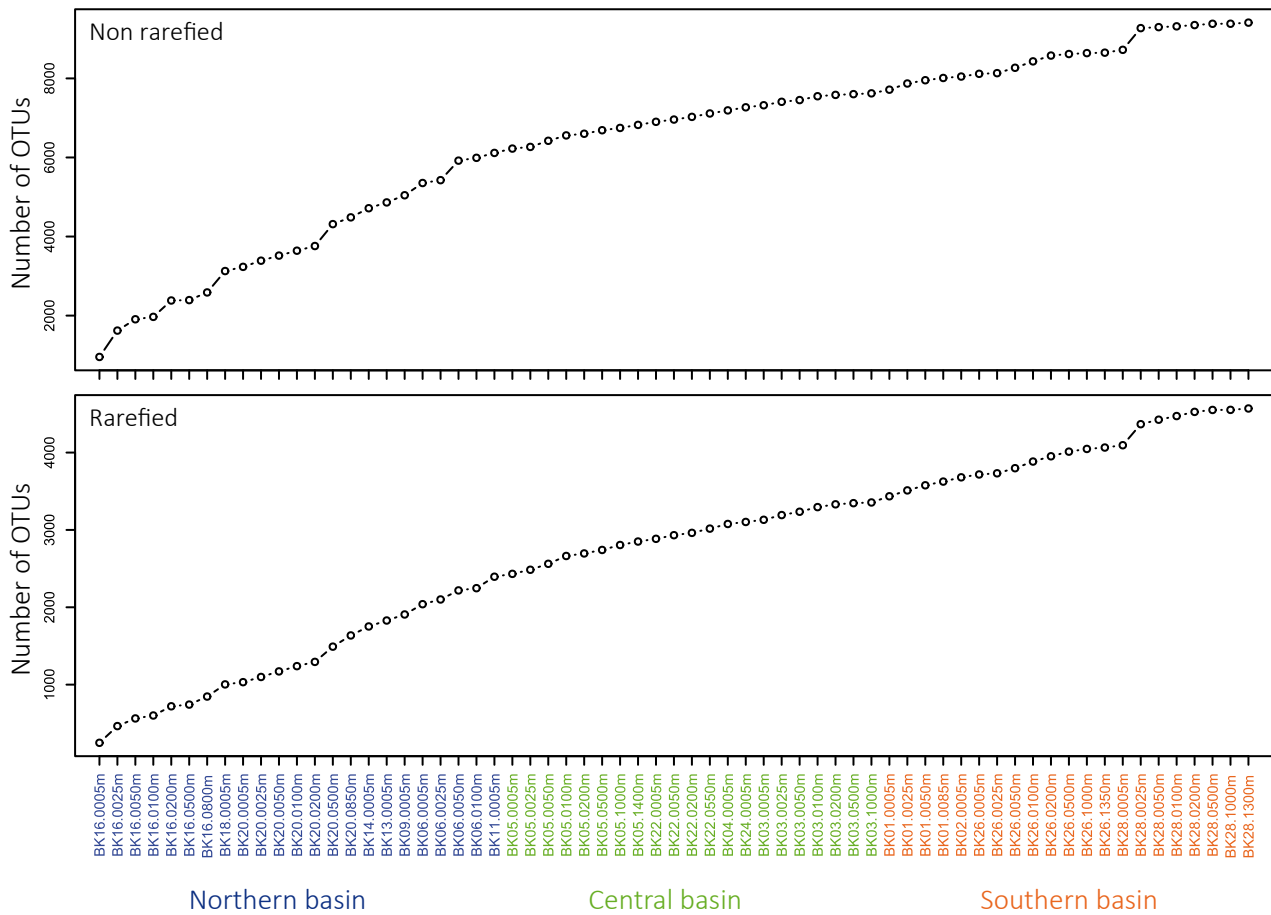


Figure S2. Accumulation curves for OTUs identified in Lake Baikal plankton samples before and after rarefaction.

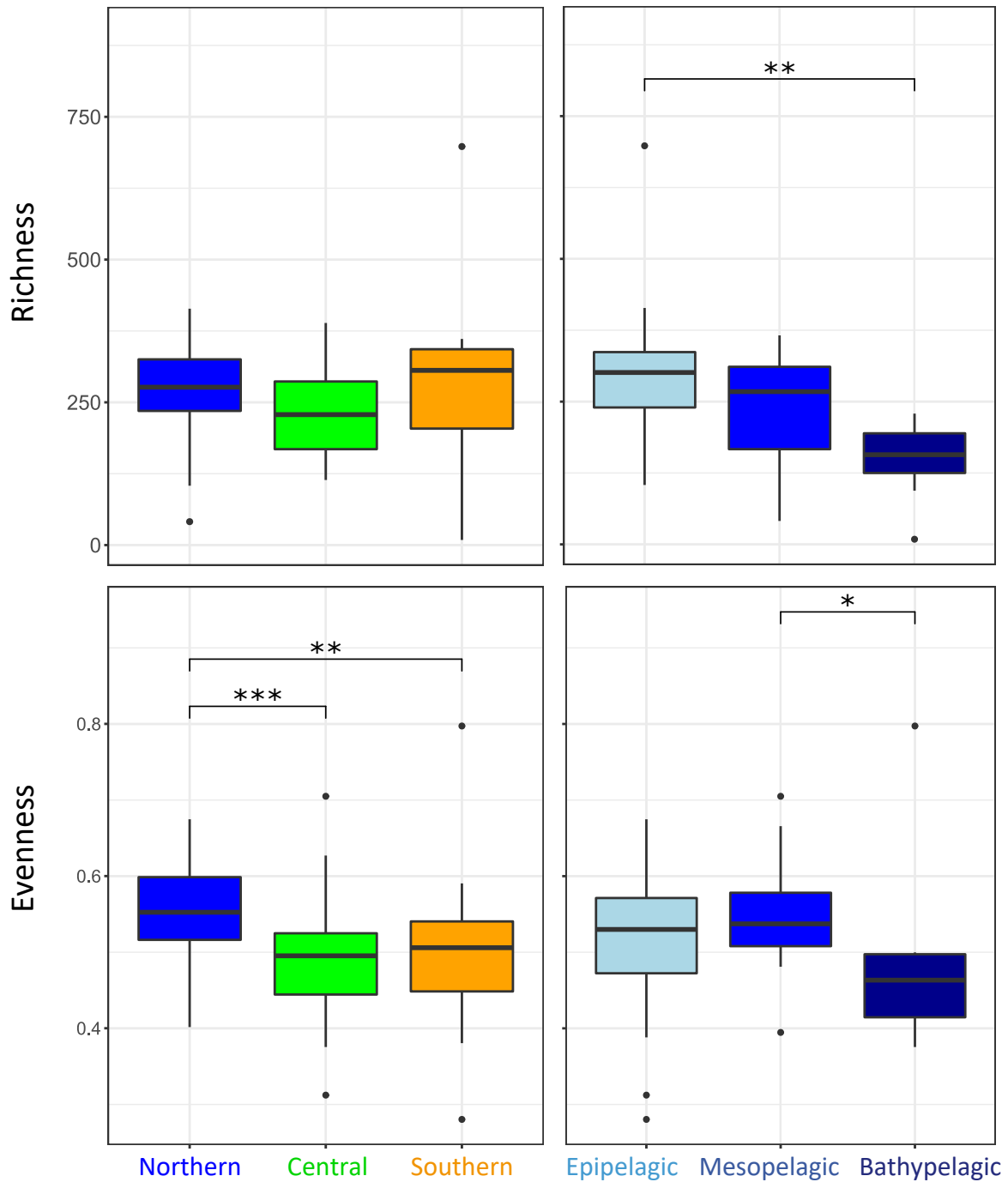


Figure S3. Box plots showing diversity (richness) and evenness values in Lake Baikal water samples as a function of basin and depth. Richness and evenness were calculated on OTUs defined at 95% 18S rRNA gene sequence identity (~genus level). Significant differences between pairs of samples are indicated (p-value ≤ 0.05 *; p-value < 0.001 ***).

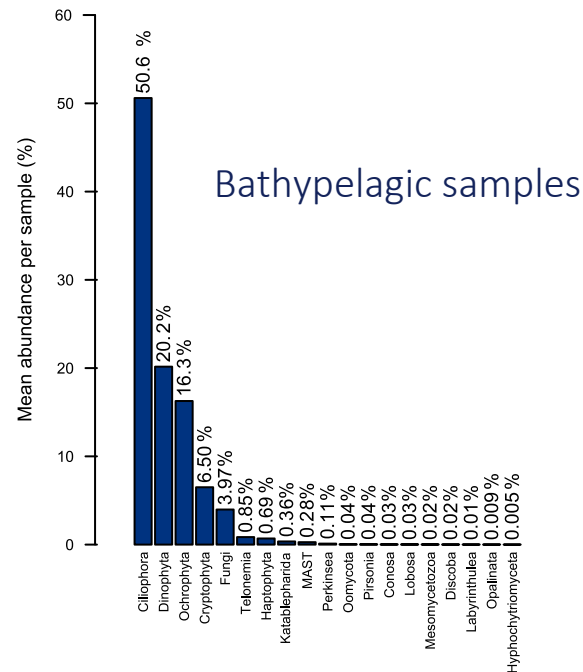
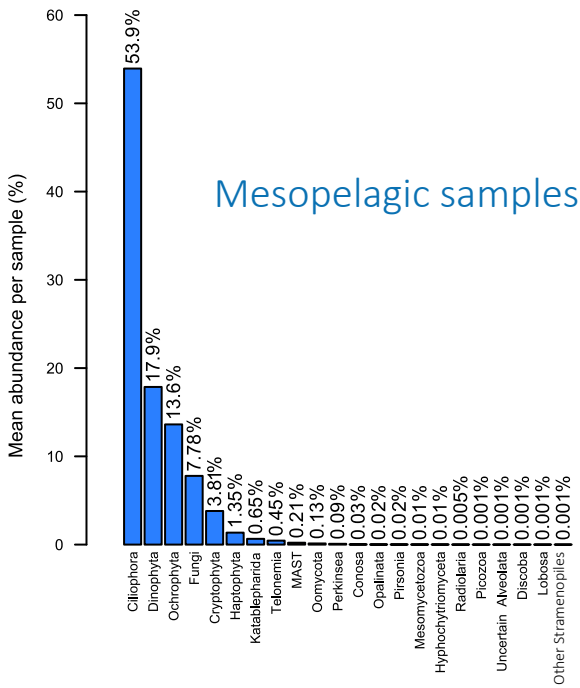
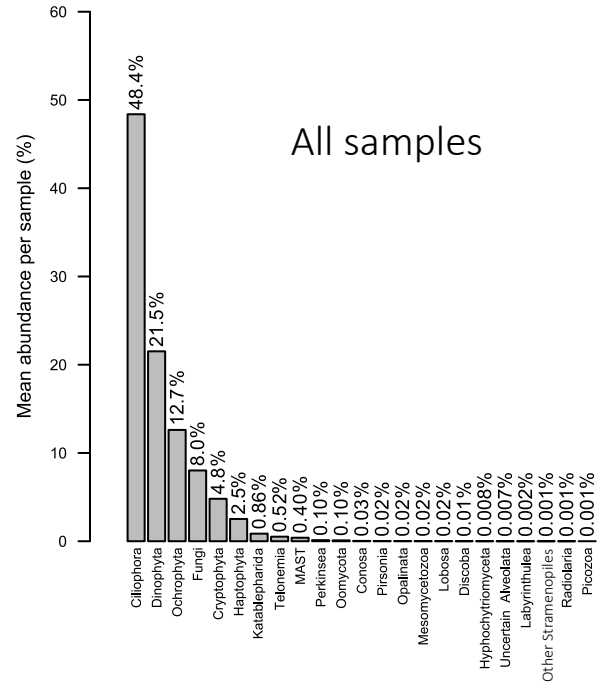
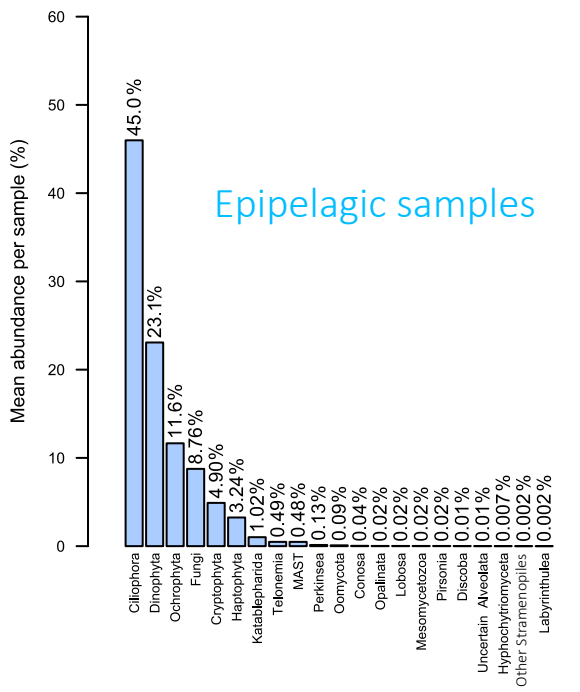


Figure S4. Rank:abundance curves of protist OTUs in Lake Baikal plankton samples. Rank:abundance curves are presented globally for the lake and by lake depth category.

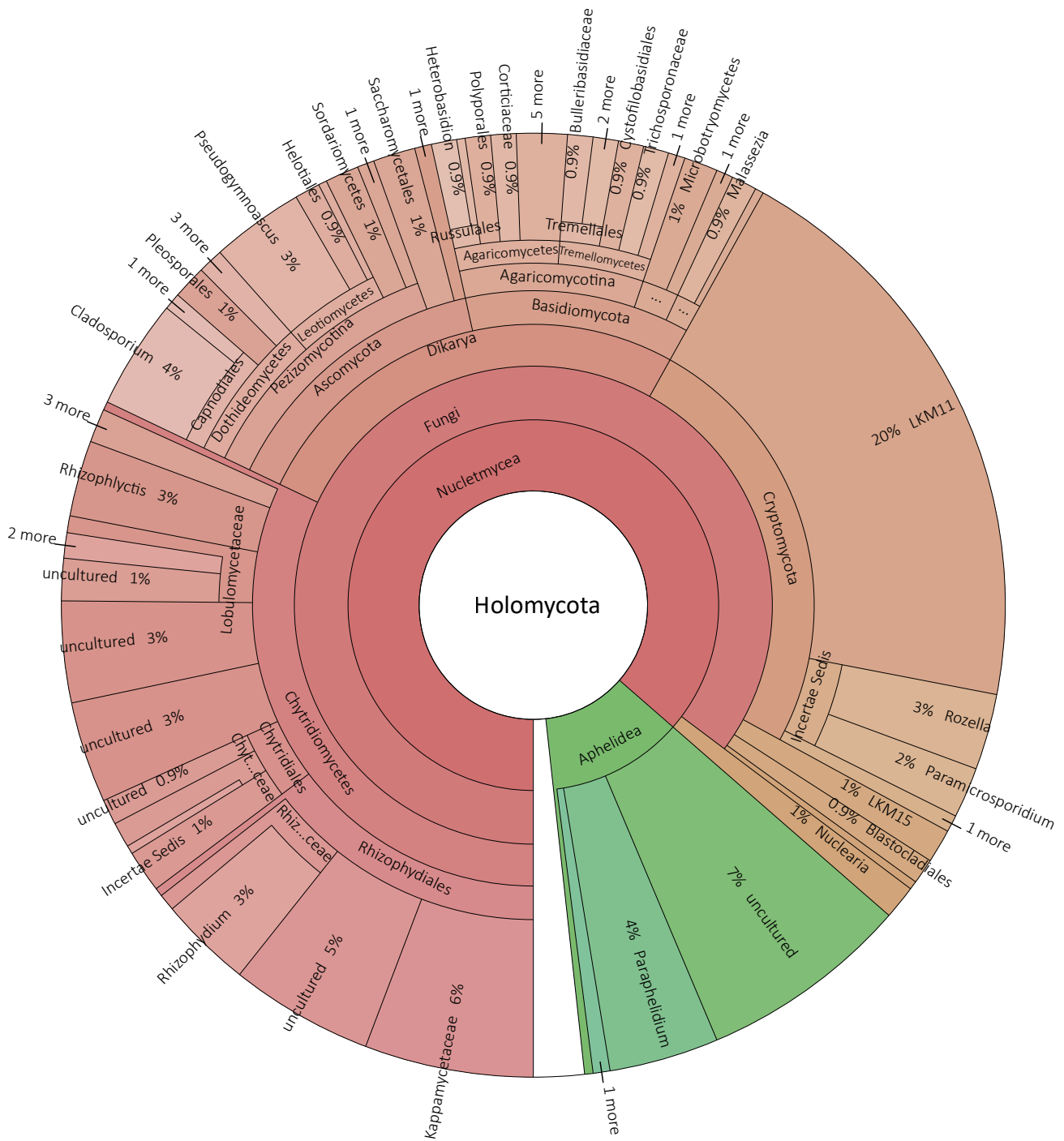


Figure S5. KRONA representation of the global diversity of Holomycota in Lake Baikal plankton. Holomycota is one of the two branches of Opisthokonta, including Fungi and related lineages (Cryptomycota, Aphelida, Nuclearida). The graph was produced using the SILVA classifier (<https://www.arb-silva.de>).

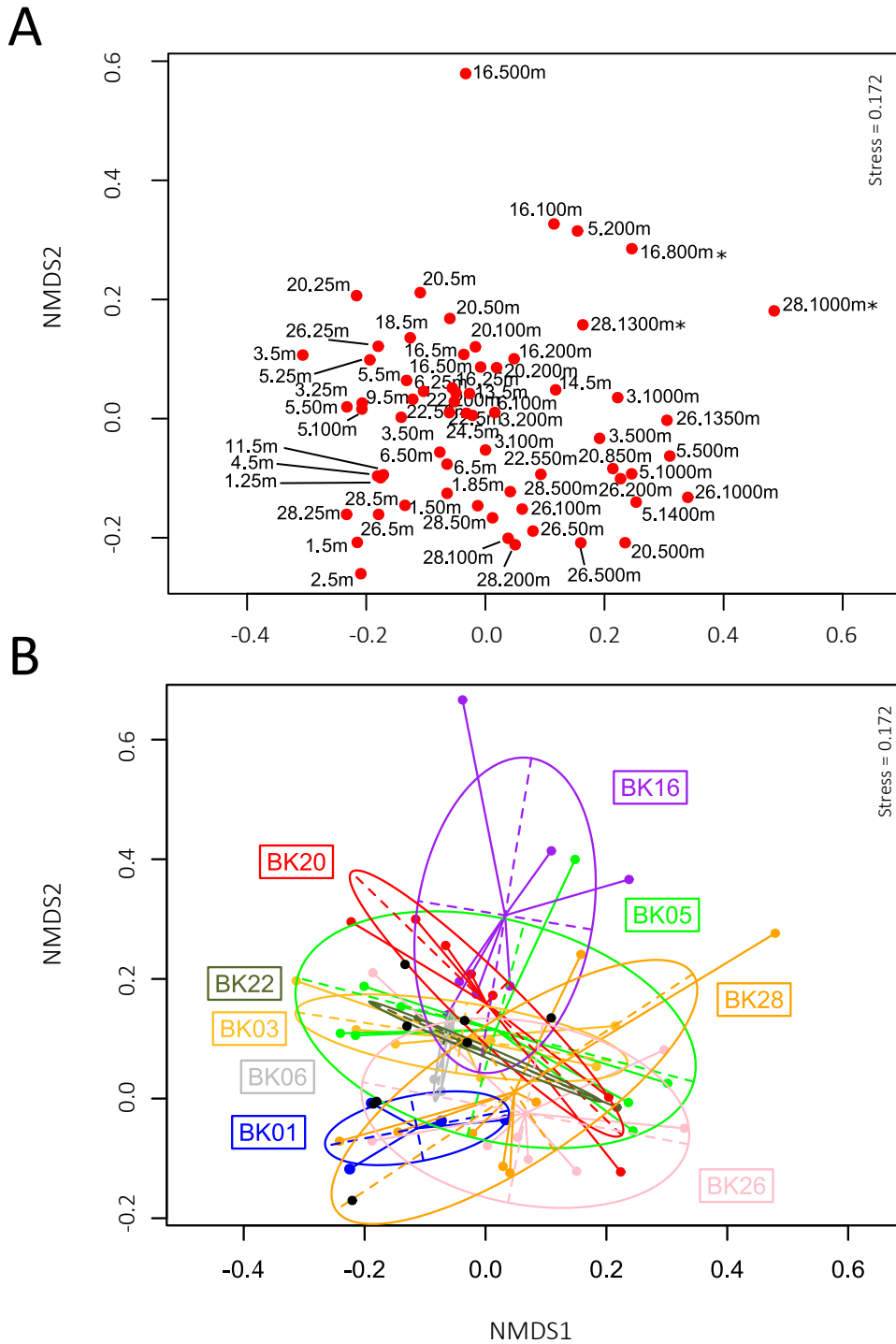


Figure S6. Non-metric multidimensional scaling (NMDS) analysis of Lake Baikal plankton samples based on differences in protist community composition. Community structure differences are based on Bray-Curtis distances established upon the distribution and relative frequency of OTUs. A, NMDS showing all sampling points. B, NMDS on the same points taking into account different water column samples. Black dots correspond to surface (5 m deep), coastal samples.

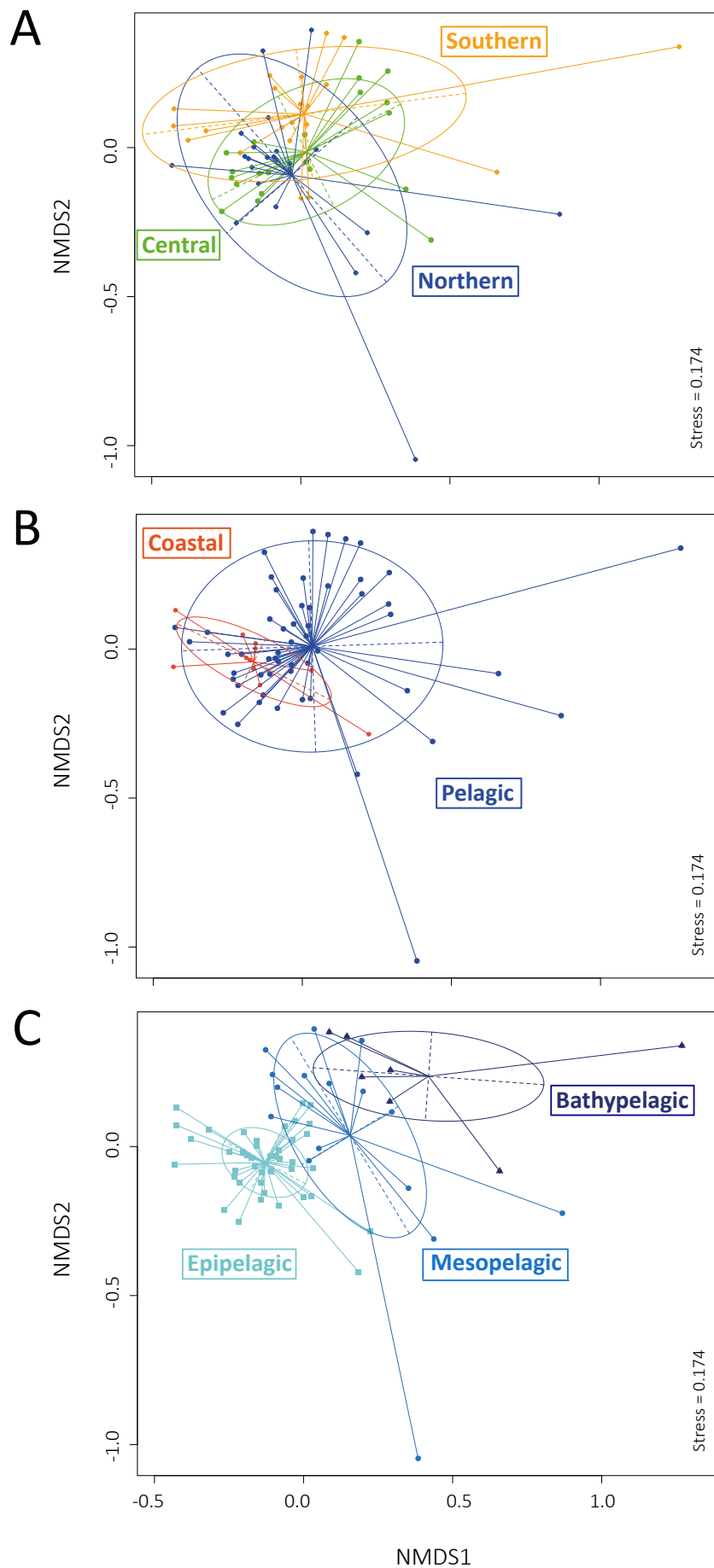


Figure S7. Non-metric multidimensional scaling (NMDS) analysis of Lake Baikal plankton samples as a function of protist community similarities based on SWARM-determined OTUs. The NMDS plot was constructed with Wisconsin-standardized Bray-Curtis dissimilarities between all samples. A, plankton samples highlighted by basin origin. B, plankton samples from coastal, shallow sites versus open water sites. C, samples grouped according to their depth origin in the water column; epipelagic (<200 m), mesopelagic (200-1000 m), bathypelagic (>1000 m).

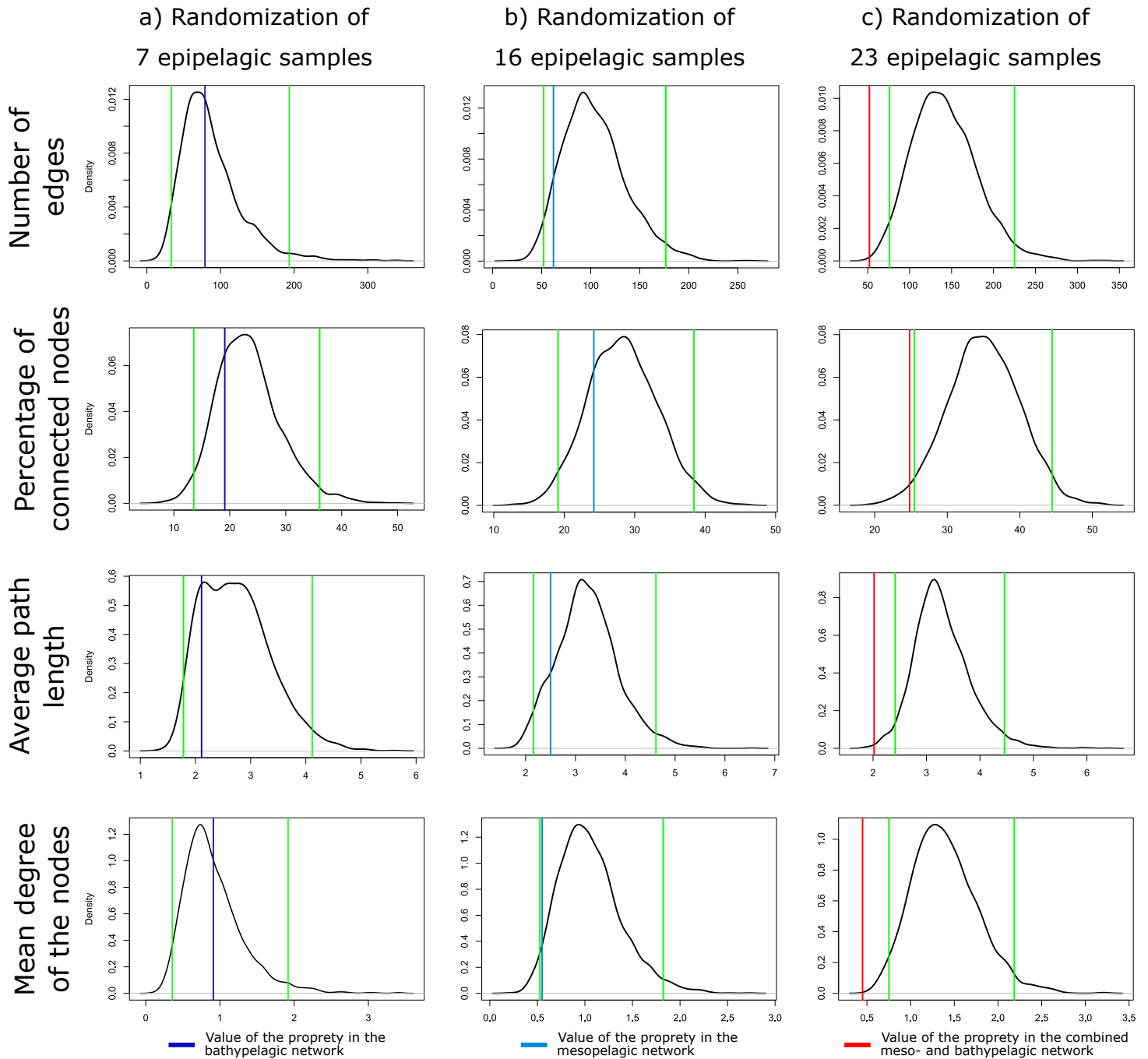
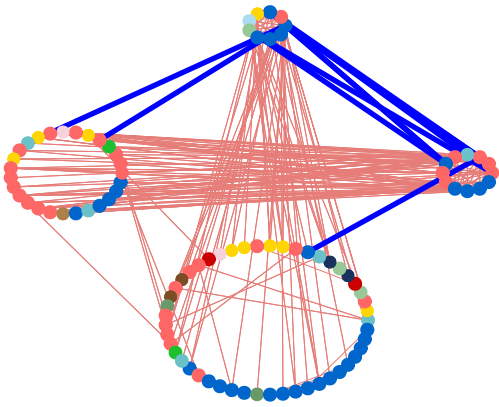
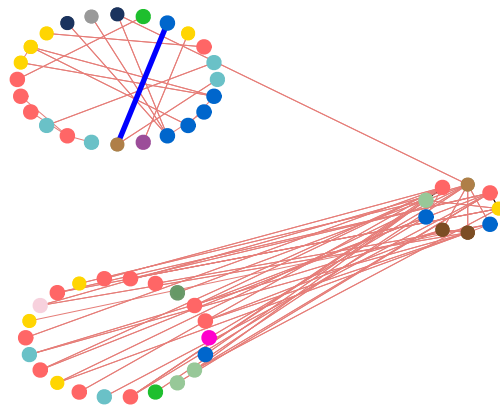


Figure S8. Bootstrap analysis on network properties, with a) Distribution of the properties of 2,500 networks constructed with 7 random samples selected among the 42 epipelagic samples. The green lines indicate the boundaries of the pvalue = 0.05. The dark blue line indicates the values of the bathypelagic network properties; b) Distribution of the properties of 2,500 networks constructed with 16 random samples selected among the 42 epipelagic samples. The green lines indicate the boundaries of the pvalue = 0.05. The medium blue line indicates the values of the mesopelagic network properties; c) Distribution of the properties of 2,500 networks constructed with 23 random samples selected among the 42 epipelagic samples. The green lines indicate the boundaries of the pvalue = 0.05. The red line indicates the values of the combined meso- and bathypelagic network properties.

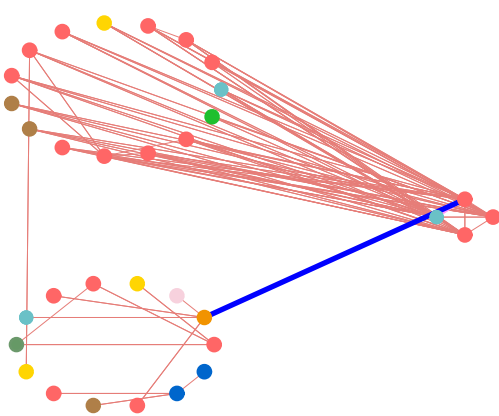
A Epipelagic



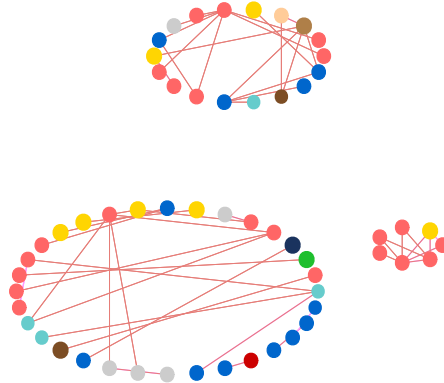
B Mesopelagic



C Bathypelagic



D Meso + bathypelagic



— Positive edge
— Negative edge

- Ciliophora
- Dinophyta
- Perkinsea
- Chlorophyta
- Diplonemida
- Cryptophyta
- Haptophyta
- Katablepharidophyta
- Telonemia
- Choanoflagellida
- Fungi
- Rozellida/Cryptomycota
- Aphelida
- Cercozoa
- MAST
- Ochrophyta
- Other Stramenopiles

Figure S9. Block-model representation of the networks of the planktonic protists in Lake Baikal for major defined depth zones. Networks were built on OTUs present in more than 20% samples and having a relative abundance higher than 0.01%. OTUs are represented by nodes and direct covariations between them, by edges.

Table S2. Identification, phylogenetic affinity and abundance of eukaryotic OTUs identified in Lake Baikal plankton.

(too large to be displayed in pdf; available in excel format)

Table S3. Diversity, abundance and distribution of eukaryotic lineages previously thought to be exclusively marine identified in Lake Baikal plankton.

Phylogenetic group	Number of OTUs	Number of reads	No. samples where they occur
Bolidophyceae	85	9442	58
Diplonema	4	505	21
Radiolaria	2	22	1
Syndiniales	73	1098	58
All MAST	95	23722	61
Specific MAST clades			
MAST1	1	2	1
MAST2	22	8250	54
MAST3	13	1296	41
MAST4	6	219	10
MAST6	4	3453	50
MAST8	1	4	3
MAST12	47	10487	59
MAST20	1	11	1

Table S4. ANOSIM analyses between pairs of Lake Baikal plankton sample groups defined as a function of depth, basin of origin and coastal vs. pelagic location. ANOSIM were calculated upon 999 permutations between pairs of sample groups.

	R	p-value
Epi-Meso	0.3821	0.0001
Epi-Bathy	0.7262	0.0001
Meso-Bathy	0.7262	0.0001
Southern-Central	0.09237	0.0001
Southern-North	0.2201	0.0001
Central-North	0.1442	0.0001
Coastal-Pelagic	0.05955	0.2339

Table S5. Tentative classification of phylogenetic lineages within broad functional categories.

Phylum	Super_Group	Lifestyle
Chlorophyta	Archaeplastida	Autotrophs
Cryptophyta	Hacrobia	Autotrophs
Dinophyta	Alveolata	Autotrophs
Haptophyta	Hacrobia	Autotrophs
Ochrophyta	Stramenopiles	Autotrophs
Rhodophyta	Archaeplastida	Autotrophs
Charophyta	Archaeplastida	Autotrophs
Apusozoa	Apusozoa	Free-living heterotrophs
Bicoecea	Stramenopiles	Free-living heterotrophs
Breviatea	Amoebozoa	Free-living heterotrophs
Centrohelioczoa	Hacrobia	Free-living heterotrophs
Cercozoa	Rhizaria	Free-living heterotrophs
Choanoflagellida	Opisthokonta	Free-living heterotrophs
Ciliophora	Alveolata	Free-living heterotrophs
Conosa	Amoebozoa	Free-living heterotrophs
Discoba	Excavata	Free-living heterotrophs
Fungi	Opisthokonta	Free-living heterotrophs
Katablepharidophyta	Hacrobia	Free-living heterotrophs
Labyrinthulea	Stramenopiles	Free-living heterotrophs
Lobosa	Amoebozoa	Free-living heterotrophs
MAST	Stramenopiles	Free-living heterotrophs
Metamonada	Excavata	Free-living heterotrophs
Radiolaria	Rhizaria	Free-living heterotrophs
Telonemia	Hacrobia	Free-living heterotrophs
Picozoa	Uncertain	Free-living heterotrophs
Mesomycetozoa	Opisthokonta	Free-living heterotrophs
Pirsonia	Stramenopiles	Free-living heterotrophs
Syndiniales	Alveolata	Putative parasites
Cryptomycota	Opisthokonta	Putative parasites
Hyphochytriomyceta	Stramenopiles	Putative parasites
Oomycota	Stramenopiles	Putative parasites
Apicomplexa	Alveolata	Putative parasites
Opalinata	Stramenopiles	Putative parasites
Perkinsea	Alveolata	Putative parasites
Aphelida	Opisthokonta	Putative parasites
Uncertain_Alveolata	Alveolata	Uncertain
Uncertain_Amoebzoa	Amoebozoa	Uncertain
Uncertain_Opisthokonta	Opisthokonta	Uncertain
Uncertain_Stramenopiles	Stramenopiles	Uncertain

Table S7. Properties of networks built upon Lake Baikal plankton OTUs for the three depth categories of the water column. Only OTUs that were present in more than 20% of samples with a relative abundance higher than 0.01% were considered.

Parameter	Epipelagic		Mesopelagic		Bathypelagic		Meso- and bathypelagic	
	Basin	%	Basin	%	Basin	%	Basin	%
Positive edges	10	4.8	1	1.61	1	1.26	0	0
Negative edges	198	95.19	61	98.38	78	98.73	52	100
Total number of edges	208		62		79		52	
Total number of nodes	206		227		173		230	
Connected nodes	97	47.08	55	24.22	33	19.07	57	24.8
Connectance	0.04		0.041		0.14		0.033	
Clustering coefficient	0.31		0.25		0.43		0.26	
Most connected node	X406443 (27)		X403371 (9)		X401791, X403173 (18)		X404559 (7)	
Average path length	3.51		2.5		2.11		2.02	
Mean node degree	2.02		0.55		0.91		0.45	