



OPEN Taxonomic composition, community structure and molecular novelty of microeukaryotes in a temperate oligomesotrophic lake as revealed by metabarcoding

Konstantina Mitsi^{1✉}, Daniel J. Richter¹, Alicia S. Arroyo¹, David López-Escardó², Meritxell Antó¹, Antonio Guillén Oterino³ & Iñaki Ruiz-Trillo^{1,4}

Microbial eukaryotes are diverse and ecologically important organisms, yet sampling constraints have hindered the understanding of their distribution and diversity in freshwater ecosystems. Metabarcoding has provided a powerful complement to traditional limnological studies, revealing an unprecedented diversity of protists in freshwater environments. Here, we aim to expand our knowledge of the ecology and diversity of protists in lacustrine ecosystems by targeting the V4 hypervariable region of the 18S rRNA gene in water column, sediment and biofilm samples collected from Sanabria Lake (Spain) and surrounding freshwater ecosystems. Sanabria is a temperate lake, which are relatively understudied by metabarcoding in comparison to alpine and polar lakes. The phylogenetic diversity of microbial eukaryotes detected in Sanabria spans all currently recognized eukaryotic supergroups, with Stramenopiles being the most abundant and diverse supergroup in all sampling sites. Parasitic microeukaryotes account for 21% of the total protist ASVs identified in our study and were dominated by Chytridiomycota, both in terms of richness and abundance, in all sampling sites. Sediments, biofilms and water column samples harbour distinct microbial communities. Phylogenetic placement of poorly assigned and abundant ASVs indicates molecular novelty inside Rhodophyta, Bigyra, early-branching Nucleotmycea and Apusomonadida. In addition, we report the first freshwater incidence of the previously exclusively marine genera *Abeoforma* and *Sphaeroforma*. Our results contribute to a deeper understanding of microeukaryotic communities in freshwater ecosystems, and provide the first molecular reference for future biomonitoring surveys in Sanabria Lake.

The tree of eukaryotes is an ideal playground for biodiversity explorers. Although land plants, animals and fungi initially attracted most of researchers' attention, the advent of molecular techniques in biodiversity assessment has revealed an enormous diversity of microbial eukaryotes outside these three groups^{1,2}. The paraphyletic assemblage of microbial eukaryotes is collectively referred to as protists³. Protists are valuable from an evolutionary perspective because by studying their life traits we gain insights into the evolutionary processes that shaped the extant eukaryotic tree of life⁴⁻⁸. In addition, protists are abundant, diverse and widespread organisms with key roles in important ecosystemic functions⁹⁻¹¹. However, despite their importance in different ecosystems as producers¹², grazers^{13,14}, predators¹⁵ and parasites¹⁶, they have attracted less attention in comparison to their prokaryotic counterparts in environmental surveys¹⁰.

Sampling based on morphological identification combined with environmental DNA (eDNA) analyses¹⁷ have shown that protists are everywhere¹⁸. However, they are not everywhere equally studied. Microbial eukaryotes have been widely explored in marine ecosystems¹⁹⁻²⁷, whereas there are fewer studies available regarding

¹Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de La Barceloneta, 37-49, 08033 Barcelona, Spain. ²Institut de Ciències del Mar (CSIC), Passeig Marítim de La Barceloneta, 37-49, 08033 Barcelona, Spain. ³I.E.S. "Escultor Daniel", C/ Gonzalo de Berceo, 49, 26005 Logroño, La Rioja, Spain. ⁴Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys, 23, 08010 Barcelona, Spain. ✉email: kcmitsi@gmail.com

their distribution and diversity in soils^{16,28–30} and in freshwater systems^{31,32}. Freshwater ecosystems are more fragmented and isolated^{33,34} in comparison to the open ocean where microbial communities are transported by currents on a global scale^{35,36}. This intrinsic lower connectivity of freshwater ecosystems hinders the dispersal of freshwater organisms and increases the genetic diversity³⁷.

Among freshwater habitats, lakes are undoubtedly the environments with the greatest number of molecular studies available^{38–40}. High mountain lakes^{40–43} and polar lakes^{44,45} have been extensively studied due to their extreme conditions of temperature, nutrient availability and UV radiation. These systems have been shown to harbour a high proportion of unclassified sequences within numerous eukaryotic lineages. Fewer molecular biodiversity surveys, however, have been conducted in lakes of temperate areas^{46,47}.

In this study, we explore the diversity of microbial eukaryotes in Sanabria Lake (Spain), a temperate lake at an altitude of 1004.1 m above sea level. Sanabria is an oligotrophic to oligomesotrophic, warm, monomictic lake with winter circulation and summer stratification. In comparison to lakes of other trophic states, oligomesotrophic lakes harbour the richest and most diverse lentic organismal communities⁴⁶. Sanabria Lake is the biggest natural glacial lake in the Iberian Peninsula⁴⁸ and has already been the subject of many traditional limnological studies^{49–56}. However, no molecular data are currently available for this freshwater system.

The aim of this study is to explore the eukaryotic microbial community of Sanabria Lake using a massively parallel sequencing approach. To do so, we targeted the V4 hypervariable region of the 18S rDNA gene. We explored the taxonomic composition of the microbial eukaryotes present in the lake and the surrounding freshwater systems, including an exploration of the protist parasite diversity. We also assessed the community structure and the compositional heterogeneity across different sampling sites, habitats, and filter size fractions. Finally, we analysed our samples using a phylogenetic placement approach to quantify molecular novelty and we identified the branches of the eukaryotic tree that harbour potentially novel abundant taxa. Sanabria Lake is a protected biotope and it is under continuous monitoring. This is the first biodiversity study of Sanabria Lake that is based on molecular data, which will constitute a reference for future monitoring efforts.

Materials and methods

Study area. Sanabria Lake is situated in the NW of Spain (42° 7' 30" N, 06° 3' 00" W) between the provinces of León and Zamora at an altitude of 1004.1 m above sea level. It was formed by glacial erosion after the Würm glaciation in the Pleistocene, and it is the only lake formed by a terminal moraine in the Iberian Peninsula⁵⁷. Sanabria Lake belongs to the Duero River Basin that has a total drainage surface of 127.3 km²⁵⁸ and its main tributary is the Tera River. The surface of the lake is 3.46 km²⁵⁸. It is divided longitudinally into two basins, one in the west with maximum depth of 46 m and another in the east with maximum depth of 51 m⁴⁸. The shoreline length is 9518 m and the maximum width is observed in the eastern basin (1530 m)⁴⁸. Regarding its mixing characteristics, Sanabria Lake is a warm, monomictic, holomictic lake⁵⁶. The mixing period extends from late November to early March, when a thermocline normally appears⁵⁶. No anoxic conditions have been observed in any layer of the water column during the thermal stratification^{54,56,58}. Sanabria Lake is considered as oligotrophic to oligomesotrophic in view of its low levels of chlorophyll *a*, nutrient concentration, phytoplanktonic biovolume values and production rates^{50,54–56,58}. The oligotrophic state of the lake is a result of its geology. Its drainage basin runs over an acid rock substrate (gneiss and granodiorites) of low solubility, making the water very poor in salts⁵⁹. The lake is part of the Sanabria Lake Natural Park (BOE 1978), a protected area that supports a population of 2 small villages (~200 residents), one in the north and the other in the west side of the lake. During the summer, the National Park receives a high influx of tourists and there are three camping sites, all located on the east side of the lake. Since 2012, the Duero International Biological station has raised concerns that Sanabria Lake is undergoing an eutrophication process due to contamination from a deficient sewage depuration system⁶⁰. However, studies based on pigment measurements and microscopy observation of the phytoplankton community do not support the eutrophication scenario and confirm the current oligotrophic state of the lake^{54,55}.

Sampling. Sampling was conducted at the beginning of the thermal stratification in March 2016. This time point was chosen because the physicochemical conditions of the lake are homogeneous after the winter mixing period, and it was expected that microbial eukaryotes would be homogeneously distributed across the lake, which would increase sampling efficiency. In addition, this time of year has the least anthropogenic impact, so any disturbance detected would indicate a permanent change rather than a temporal variation due to the presence of a stressor.

To explore the diversity and the relative abundance of microbial eukaryotes in Sanabria Lake, 82 samples of water, sediment and biofilms from ten different locations were collected. Water samples were collected from five sampling sites in the lake basin (S1–S5), a tributary stream (S6–S7) and a nearby pond (S8–S10) (Fig. 1, Supplementary Table 1). A tributary is a stream or river that flows into a larger water body such as another river or a lake. A total of six different habitats were included in the sampling design: (i) two pelagic sampling sites in the lake, one in the west basin (S1) and another in the east basin (S4), (ii) the mouth of the River Tera (S2), which experiences the greatest anthropogenic impact in the studied ecosystem, (iii) a coastal area in the lake near two islets (Islas Moras) (S3), (iv) sulphurous waters in the east basin (S5), (v) water from a tributary stream (S6–S7) and (vi) water from three nearshore sites in a nearby pond (S8–S10). Water samples were collected from each of these habitats and were taken from the surface, the Deep Chlorophyll Maximum (DCM) and the deepest point above the sediment. The standard variables (turbidity, temperature, fluorescence) in the lake's main water body (S1–S5) were measured using a CTD SD204 (SAIV A/S) device. The water samples were prefiltered using filters of 2000 µm and 200 µm to remove debris and large multicellular organisms, and the size fractions above 200 µm were discarded and not included in the study. The water was then filtered sequentially using filters of 20 µm, 5 µm and 0.8 µm targeting the microplankton (20–200 µm), the nanoplankton (5–20 µm) and the pico/

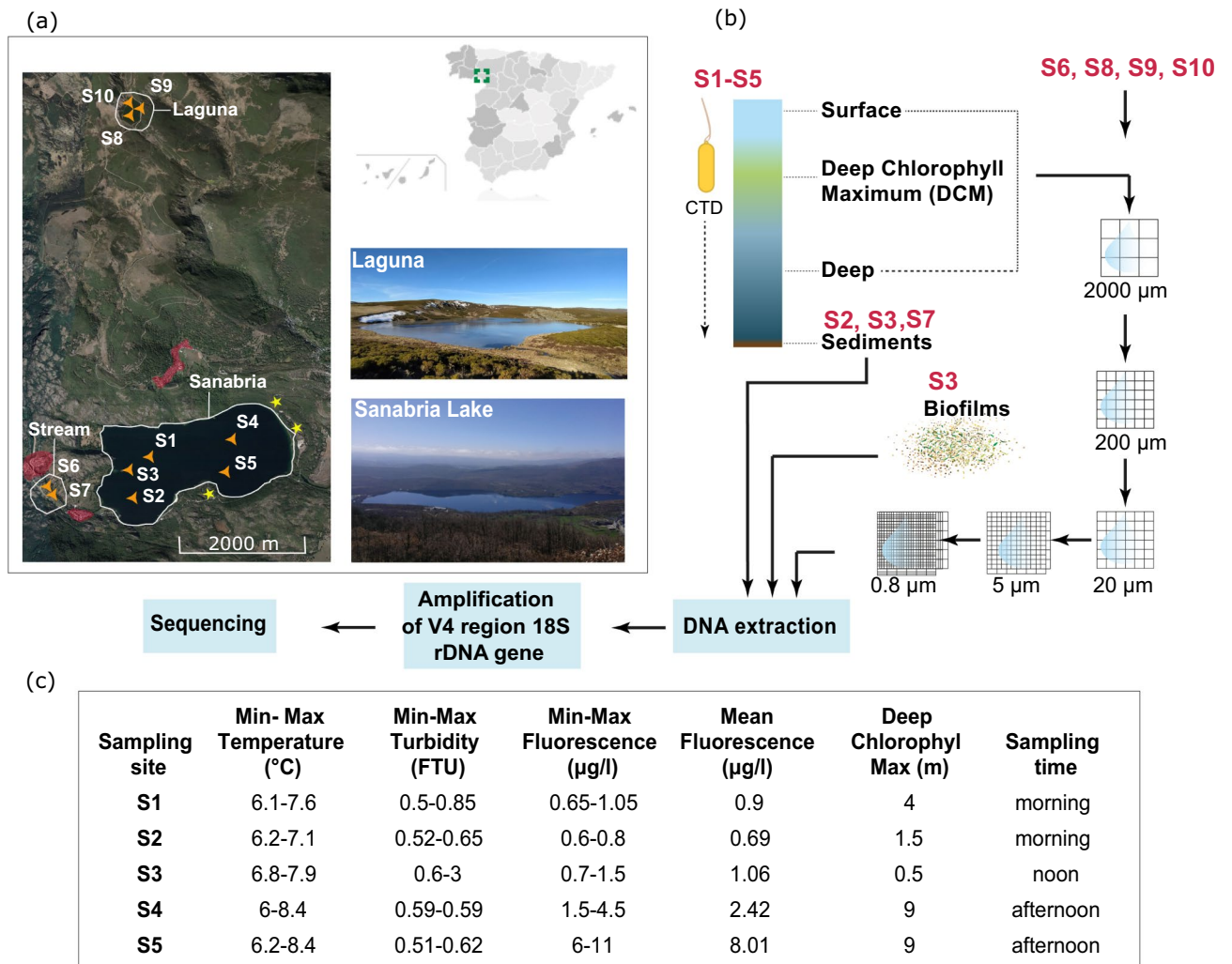


Figure 1. Sampling information. (a) The map on the top right shows the position of Sanabria Lake in the Iberian Peninsula. The satellite map (Source: Google Earth web, Imagery date: 12/07/2019, Camera: 16 km, Coordinates: 42°08'37"N 6°43'257"W) shows the study area. The sampling sites are pointed by orange triangles. Nearby villages are delimited by red coloured areas. Camping sites are pointed by yellow stars. (b) The sampling protocol. Filters of 2000 µm and 200 µm contained mainly multicellular organisms and were discarded. Sample S6 is water and S7 is sediment from an upstream tributary stream. Samples S8–S10 are water samples from a nearby small pond (Laguna de los Peces) that is not connected to the main water body. (c) CTD data collected in Sanabria Lake (sites S1–S5) during sampling.

nanoplankton (0.8–5 µm) communities, respectively. In addition to water samples, sediments were collected from three sampling sites (S2, S3, S7) and 12 epilithic biofilm samples were collected from one sampling site (S3). All samples were placed in 2 ml cryovials and stored at –80 °C until DNA extraction.

DNA extraction, PCR amplification and sequencing. The filters were chopped using sterile scissors, and the sediments and biofilms were homogenised before further processing. Whole genomic DNA was extracted using the standard protocol of the PowerSoil DNA Isolation Kit (MO BIO). The V4 hypervariable region of the 18S rRNA gene was amplified using the universal eukaryotic V4 primers TAReuk454FWD1 (5'-CCAGCA(G/C)C(CT)GCGGTAATTCC-3') and TAReukREV3 (5'-ACTTTCGTTCTTGAT(CT)(A/G)A-3')²⁵. The amplicons were sequenced using the Illumina MiSeq platform at Centro Nacional de Análisis Genómico (CNAG, Barcelona, Spain). The sequences were demultiplexed and the barcodes were trimmed by the sequencing facility. The raw sequencing data were submitted to the European Nucleotide Archive (ENA) under the accession number PRJEB23911 (ERP105690).

Amplicon Sequence Variants (ASVs) analysis. The raw reads were analysed following a clustering-free approach. The Illumina demultiplexed paired-end sequenced dataset was processed using the R package dada2⁶¹. The read quality profiles were visualised using the function plotQualityProfile. The quality of the forward and reverse reads started decreasing after the position 250 and 240, respectively. The function filterAndTrim was

used to discard low quality sequences using standard filtering parameters (maxN = 0, maxEE = c(2,2), truncQ = 2, rm.phix = TRUE, compress = FALSE, multithread = TRUE) and to trim the forward reads in the position 250 and the reverse reads in the position 240. The error model was calculated from our data with the function learnErrors and the estimated error rates were visualised with the function plotErrors. The reads were dereplicated using the function derepFastq and sequence-variants from our dereplicated data were inferred using the function dada. The forward and the reverse reads were merged to obtain the full denoised sequences using the default 12 bases overlap region and the paired reads that did not exactly overlap were removed. The amplicon sequence variant table (ASV) table was constructed using makeSequenceTable, then chimeras were removed with removeBimeraDenovo and finally the number of reads that made it through each step of our analysis was inspected (Supplementary Table 2). Taxonomy was assigned with the function assignTaxonomy that uses the Ribosomal Database Project (RDP) classifier together with the Protist Ribosomal Reference database (PR2) (v. 4.12.0)⁶² formatted for DADA2 (<https://github.com/pr2database/pr2database/releases>). An ASVs table that contains a total of 31,225 ASVs, the ASV counts per sample, and their taxonomy was generated using Biostrings::DNAStringSet and Biostrings::writeXStringSet from the Biostrings (version 2.58.0) R package in Bioconductor.

Statistical analyses. The taxonomy, abundance, and metadata were combined to generate a phyloseq object⁶³ (Supplementary Table 3). Different phyloseq datasets were generated by subsetting the initial phyloseq object (Supplementary Table 3) using the commands phyloseq::subset_samples and phyloseq::subset_taxa. Rarefaction curves were plotted using the function phyloseq::rarecurve to explore whether all included samples had reached saturation. Samples that did not reach saturation were removed. The phyloseq⁶³, psadd (<https://rdr.io/github/cpauvert/psadd/>), and Fantaxtic (<https://rdr.io/github/gmteunisse/Fantaxtic/>) R packages were used to plot the taxonomic composition of the datasets.

Each dataset was rarefied at the minimum sample depth (Supplementary Table 3) in order to simulate even numbers of reads per sample. Rarefaction enables clustering samples according to their biological origin and permits fair comparison of diversity metrics among the samples⁶⁴. Alpha and beta diversity were calculated in the subsampled datasets. Nine different alpha diversity metrics (Observed species, Chao1, se.chao1, ACE, se.ACE, Shannon, Simpson, InvSimpson, Fisher) were calculated using the function phyloseq::estimate_richness in order to examine whether alpha diversity estimates vary depending on the metric used. The significance of the difference in species richness was tested with pairwise comparisons using the Wilcoxon rank sum test, controlling for family wise error rate with the Holm–Bonferroni method⁶⁵. Evenness was calculated according to Pielou⁶⁶ and plotted as violin plots in the ggplot2 R package⁶⁷. Beta diversity was measured using the Bray–Curtis dissimilarity statistic after converting ASVs abundances to frequencies within samples. To test the effects of habitat, sampling depth, sampling site, chlorophyll maximum and thermocline across samples, permutational multivariate analyses of variance (PERMANOVA) based on Wisconsin-standardized Bray–Curtis dissimilarities (Supplementary Table 6) was performed using the adonis function of the vegan package. Patterns of beta diversity were assessed via non-metric multidimensional scaling ordination (NMDS) also on Bray–Curtis dissimilarities using the function phyloseq::ordinate and were plotted using the function phyloseq::plot_ordination. The significance of groups revealed by NMDS was tested with analysis of similarity (ANOSIM) tests with 999 permutations (Supplementary Table 7).

Phylogenetic novelty analysis. A phylogenetic placement approach was used to explore the molecular novelty present in the dataset. When taxonomy assignment based on sequence similarity methods fail, the most reliable way to classify an unknown sequence is by phylogenetic inference. However, large datasets of short-read query sequences produced by Illumina NGS technology such as the one generated by this study can not be analysed using traditional likelihood-based phylogenetic inference methods due to high computational complexity and lack of phylogenetic signal that results in poor branch support and biases such as long-branch attraction. Phylogenetic placement was proposed as a way to overcome these limitations and to bring the inferential power offered by likelihood-based approaches to large, short-read data sets⁷⁰.

Two reference trees were built, one to encompass all eukaryotic diversity and another that was specific to encompass the extant diversity of unicellular Holozoa. The alignments for the reference trees were built with MAFFT v7.309⁶⁸ with automatically selected strategy according to data size and were trimmed with trimal v1.4.rev15⁶⁹ using the automated1 algorithm. The reference trees were constructed in RAxML⁷¹ under the GTR-GAMMA model. Node support was estimated by 100 rapid bootstrap replicates. The query sequences (QS) were aligned to the reference alignments with PaPaRa (version 2.5)⁷². Phylogenetic placements were produced using RAxML-EPA⁷² and visualised with iTOL⁷³.

Results and discussion

Abiotic parameters indicate ecological disturbance in the east basin of Sanabria Lake. We collected data on temperature, water turbidity, and chlorophyll *a* in order to describe the physical conditions of the lake at the time of sampling. We assessed the mixing state of the lake by temperature. Our aim was to collect the samples at the beginning of the thermal stratification when the eukaryotic microbial community is expected to be homogeneously distributed in the water body following the winter mixing. Water temperature at the surface ranged from 7.1 °C to 8.4 °C and in the deepest sampling points ranged from 6 °C to 6.85 °C, with a mean range of 1.62 °C (Fig. 1, Supplementary Table 1). These temperature measurements agree with the previously recorded temperatures during the homeothermic state of the lake that range between 4 to 7 °C⁵⁶ and confirm the mixing state of the lake.

We assessed the trophic state of Sanabria Lake based on water turbidity and chlorophyll *a* values. Water turbidity is measured in FTU (Formazin Turbidity Units) and is an indicator of the trophic state of a lake as it is

related to the concentration, type, and size of the suspended particles in the water⁷⁴. During our sampling, turbidity values in Sanabria Lake were extremely low in all the sampling sites and ranged from 0.5 to 0.85 FTU (Fig. 1). These values are comparable to those in ultra-oligotrophic alpine lakes⁷⁵. Chlorophyll *a* is a reliable indicator to assess the trophic state of a lake with high values to correspond to more eutrophic ecosystems⁷⁶. Chlorophyll *a* mean values in Sanabria Lake have increased in the last fifty years (Supplementary Table 5) but they have not exceeded the levels that characterise oligotrophic lacustrine ecosystems. Together, these measurements confirm the overall oligotrophic status of the lake at the time of sampling.

We observed that chlorophyll *a* values differed between east and west basin during our sampling (Fig. 1). In Sanabria's west basin (samples S1–S3), the mean value of chlorophyll *a* was below the reference value (1.5 µg/L). The reference value defines the equilibrium ecological state of the lake and confirms the absence of ecological disturbances. However, the mean values of chlorophyll *a* in Sanabria's east basin (samples S4–S5) exceeded the reference values indicating the presence of ecological disturbance (Fig. 1). Values of chlorophyll *a* above 4.2 µg/L are linked to a Good-Moderate ecological state and values above 7.1 µg/L are linked to a Moderate-Poor ecological state⁷⁷. Our results showed that there was some ecological disturbance that altered the values of chlorophyll *a* in the east basin of Sanabria Lake at the time of sampling. The altered values of chlorophyll *a* in the east basin may be related to higher anthropogenic impact due to the presence of three camping sites on this side of the lake. Chlorophyll *a* values measured in Sanabria's east basin in March 2017⁵⁵ are lower than the ones presented in our study, implying that the stressor was temporal and that water quality has been restored.

The V4 hypervariable region captures the microeukaryotic diversity of Sanabria Lake. To characterise the diversity of the eukaryotic community in Sanabria Lake, we sequenced the V4 hypervariable region of the 18S small subunit (SSU) rRNA gene. We chose to sequence the V4 over other hypervariable regions of the 18S rRNA gene because it provides taxonomic resolution close to that of the full-length gene⁷⁸ and it is the most suitable hypervariable region to use for phylogenetic placement¹⁶. A total of 15,947,744 reads from 82 samples were filtered, dereplicated and merged resulting in 31,225 Amplicon Sequencing Variants (ASVs). The study of multicellular organisms was out of the scope of the present work and thus most multicellular organisms were discarded by using physical filters of 2000 µm and 200 µm. However, some environmental DNA (eDNA) that originates from cellular material shed by multicellular organisms into the lake was sequenced together with the community DNA of unicellular eukaryotes. For our subsequent analyses, we bioinformatically filtered out all ASVs that were assigned to animals (Division/Class = Metazoa), land plants (Division = Streptophyta) and typical terrestrial fungi (Class/Order = Ascomycota, Class/Order = Basidiomycota) (Supplementary Table 3, dataset D3). After the removal of multicellular taxa, 27,790 microeukaryotic (protist) ASVs remained. We evaluated the sampling depth and the representation of microbial eukaryotes in our samples using rarefaction curves (Supplementary Figure 1). The curves reached a plateau for all samples, indicating that most of the microbial richness present in Sanabria Lake and the surrounding freshwater systems was successfully captured by our study.

Spatial biodiversity patterns. To evaluate the intra-sample diversity of Sanabria Lake and the surrounding water bodies, we calculated nine different alpha-diversity indices (Supplementary Table 3). To avoid potential biases in diversity estimates due to differences in the total number of reads, we randomly subsampled the ASVs to the minimum depth of our dataset (Supplementary Table 3, dataset D3, min sample depth = 31,361 reads) before calculating the alpha-diversity indices. The number of total taxa reported was not affected by subsampling. We compared the diversity of the different water body types and we found that samples collected in the tributary stream showed significantly higher intra-sample diversity (Fig. 2) and greater evenness (Supplementary Figure 9) compared to samples from Laguna (pond) and Sanabria (lake) (Wilcoxon rank sum test P value < 0.01). Previous studies have shown that small water bodies like ponds and streams can contribute significantly to regional biodiversity of macrophytes and macroinvertebrates⁷⁹. Our data support the hypothesis that the same is true for microeukaryotes. This result pinpoints the importance of small water bodies as biodiversity reservoirs and contrasts with their relative status in national monitoring and protection strategies, where they are frequently ignored. Regarding the different habitats, sediments harbour the richest microeukaryotic communities (Fig. 2). Sediments have been shown to harbour richer communities than the water column for other groups of organisms like bacteria⁸⁰ and marine diatoms⁸¹. However, we cannot exclude that part of the diversity recorded in the sediments can be attributed to either dormant stages of planktonic microeukaryotes or dead cells that were recently settled from the water column.

To test the effect of abiotic factors in the protist community structure, we carried out permutational multivariate analysis of variance (PERMANOVA) using the Bray–Curtis dissimilarities of the ASVs between communities as a function of sample spatial origin (Supplementary Table 6). All factors tested by PERMANOVA tests revealed significant differences in protist communities as a function of site (Sanabria Lake, Laguna, Stream), sampling site (S1–S10), position regarding the chlorophyll maximum (on–off), position regarding the thermocline and habitat (water column, sediments, biofilms) (Supplementary Table 6).

To visualise the compositional differences in the community structure of protists we applied non-metric multidimensional scaling (NMDS). The communities from Sanabria Lake, the tributary stream and the Laguna were clearly separated in an ordination based on sampling site (Fig. 3a). In Sanabria Lake, the habitat was the main grouping factor of the microbial community structure, resulting in clustering of the communities from the water column, the biofilms, and the sediments (Fig. 3b). Furthermore, the community of microbial eukaryotes in the water column of Sanabria Lake was clearly segregated as a function of the filter size fraction and not the sampling depth (Fig. 3b). This is what we expected given that we sampled at the beginning of the thermal stratification after the winter mixing at the point of maximum homogeneity of the community. Our observations were statistically

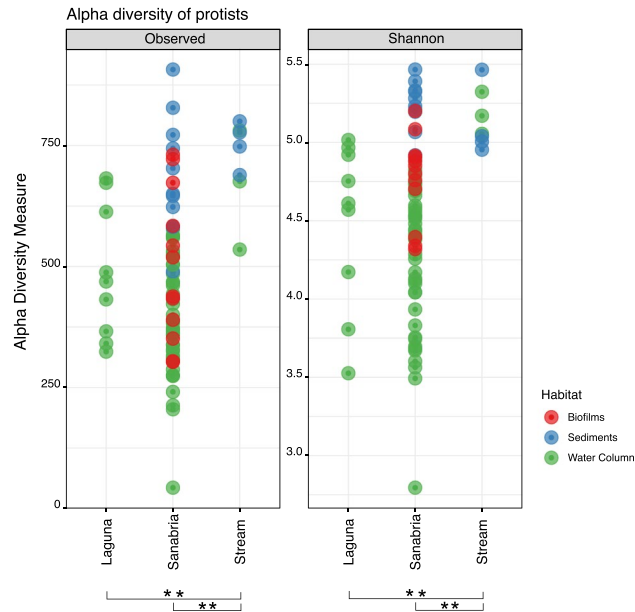


Figure 2. Alpha diversity of protists across the sampling sites. Each dot represents a sample and the colour code indicates the habitat of origin. Significant differences between pairs are indicated by double asterisks (p-value $\leq 0.01^{**}$).

supported by ANOSIM tests (Supplementary Table 7), which showed significant and marked differences among communities according to habitat, sampling site, and depth (Supplementary Table 1).

Given that chlorophyll *a* values differed between east (S1, S2, and S3 sampling sites) and west basin (S4 and S5 sampling sites) (Fig. 1), we investigated whether this difference is reflected to the structure of their microeukaryotic communities (Supplementary Figure 2). For the purpose of this analysis we only included samples from the water column, because we do not have sediments and biofilms samples from the east basin and we already know that biofilm, sediments, and water column microeukaryotic communities significantly differ between them. Permutational multivariate analysis of variance showed that east and west basin water column microeukaryotic communities do not differ significantly between them (SumsOfSqs = 0.3077, MeanSqs = 0.30768, F.Model = 0.85978, $R^2 = 0.02104$, $\text{Pr}(> F) = 0.619$).

To conclude, our results suggest that the community structure in Sanabria Lake and the surrounding freshwaters is characterised by spatial variation. The habitat is a major factor that shapes the community structure after the winter mixing period. Sediments, biofilms, and water column harbour compositionally heterogeneous microbial communities that are driven by the unique environmental parameters that characterise them.

Taxonomic composition of the protist community. To gain an overview of the microeukaryotic taxonomic composition in the Sanabria Lake and the surrounding freshwater systems, we plotted the relative abundance of ASVs at division level (based on the PR2 taxonomy) across sampling sites (Fig. 4). The phylogenetic diversity of ASVs covered all currently recognized eukaryotic supergroups^{82,83}. The group of Stramenopiles was the most abundant supergroup in all sampling sites, accounting for the 33% of total reads in Sanabria Lake, 34% in the nearby pond (Laguna) and 40% in the tributary stream respectively (Supplementary Figures 3, 4, and 5). In addition to being abundant, Stramenopiles were diverse, encompassing 22% of total ASV richness (6988 ASVs) (Supplementary Table 3). Among Stramenopiles, Ochrophyta was the most abundant group in all sampling sites (Supplementary Figure 6). Most Ochrophyta in the tributary stream (85%) and Laguna (81%) were affiliated with Chrysophyceae (Supplementary Figure 6), a group that is generally common in low-nutrient lakes⁸⁴. In Sanabria Lake, together with the Chrysophyceae (36%), we report a high relative abundance of Bacillariophyta (37%) and Synurophyceae (24%) within Ochrophyta, two phototrophic lineages that produce silica skeletons or scales (Supplementary Figure 6). Alveolata was the second most abundant and diverse supergroup, accounting for 26–28% of the total eukaryotic reads in each site (Supplementary Figures 3, 4, and 5) and a total of 4609 ASVs in the study (Supplementary Table 3).

The plankton community of Sanabria Lake (excluding the surrounding freshwater systems) was dominated by Ochrophyta (in the Stramenopiles supergroup; 26%), Ciliophora (Alveolata; 14%), Dinoflagellata (Alveolata; 10%), Cercozoa (Rhizaria; 10%), Cryptophyta (10%), and unicellular Opisthokonta (7%) (Supplementary Figure 3). The presence of all these groups except for unicellular Opisthokonta was previously documented by light microscopy in Sanabria Lake⁵⁶. We further explored the taxonomic composition of Sanabria Lake by separately examining benthic and pelagic samples. The taxonomic composition of the benthic protist community, as represented by ASVs in the sediments, was dominated by Stramenopiles (36%), Alveolata (29%), Rhizaria (13%), Opisthokonta (7%), Amoebozoa (5%), Archaeplastida (3%), Hacrobia (3%), and Excavata (3%). In contrast,

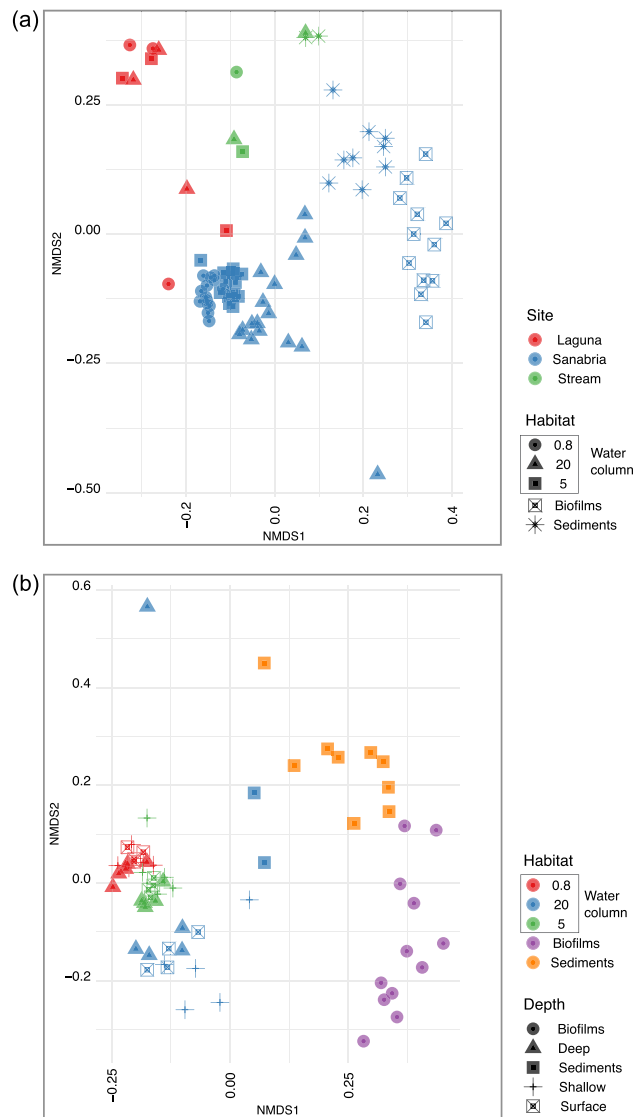


Figure 3. Beta diversity. Reduced-space NMDS plot showing microbial eukaryotes community structure based on Bray–Curtis dissimilarity. **(a)** Dissimilarity calculated from the rarefied at even depth (31,361 reads) abundances of protist ASVs in all samples (Supplementary Table 3, dataset D3). (Stress = 0.2083063, Procrustes: rmse 0.03612112 max resid 0.1863833), **(b)** Dissimilarity calculated from the abundances of rarefied at even depth (31,361 reads) ASVs present only in Sanabria samples (dataset D5) (Stress = 0.1676406, Procrustes: rmse 2.784844e−06 max resid 1.93149e−05).

the planktonic microbial community was characterised by the prevalence of Hacrobia (16%) as the third most abundant eukaryotic supergroup after Stramenopiles (31%) and Alveolata (28%). The planktonic Hacrobia⁸⁵ in Sanabria Lake included Cryptophyceae (84%), Katablepharidophyta (13%), Centrohelioczoa (1.5%), Telonemia (1%), and Haptophyta (0.5%). Excluding Cryptophyceae, this is the first record for these taxonomic groups in Sanabria Lake. (Katablepharidophyta were previously classified inside Cryptophyceae until electron microscopy and molecular phylogenies provided evidence to consider them as a separate taxonomic group⁸⁶).

Protist parasites in a temperate oligomesotrophic lake. Here we provide the first description of the taxonomic composition of the unicellular eukaryotic parasites (Supplementary Table 3, dataset D6) present in Sanabria Lake, the biggest natural lake in the Iberian Peninsula. Parasitic unicellular eukaryotes modulate large scale ecological processes by regulating the abundance and dynamics of their host population⁸⁷. As their study by microscopy is tedious, little was known about their prevalence in freshwater systems until the advent of metabarcoding⁸⁸.

The parasites accounted for 21.3% (5925) of the total protist ASVs identified in our study. The parasitic community composition was dominated in all sampling sites by Chytridiomycota, whose relative abundance within parasitic taxa was 29% in the tributary stream, 32% in Sanabria Lake and 42% in Laguna. The prevalence of Chytridiomycota in the pelagic zone of lakes has been previously reported^{89,90}. Chytridiomycota, which includes

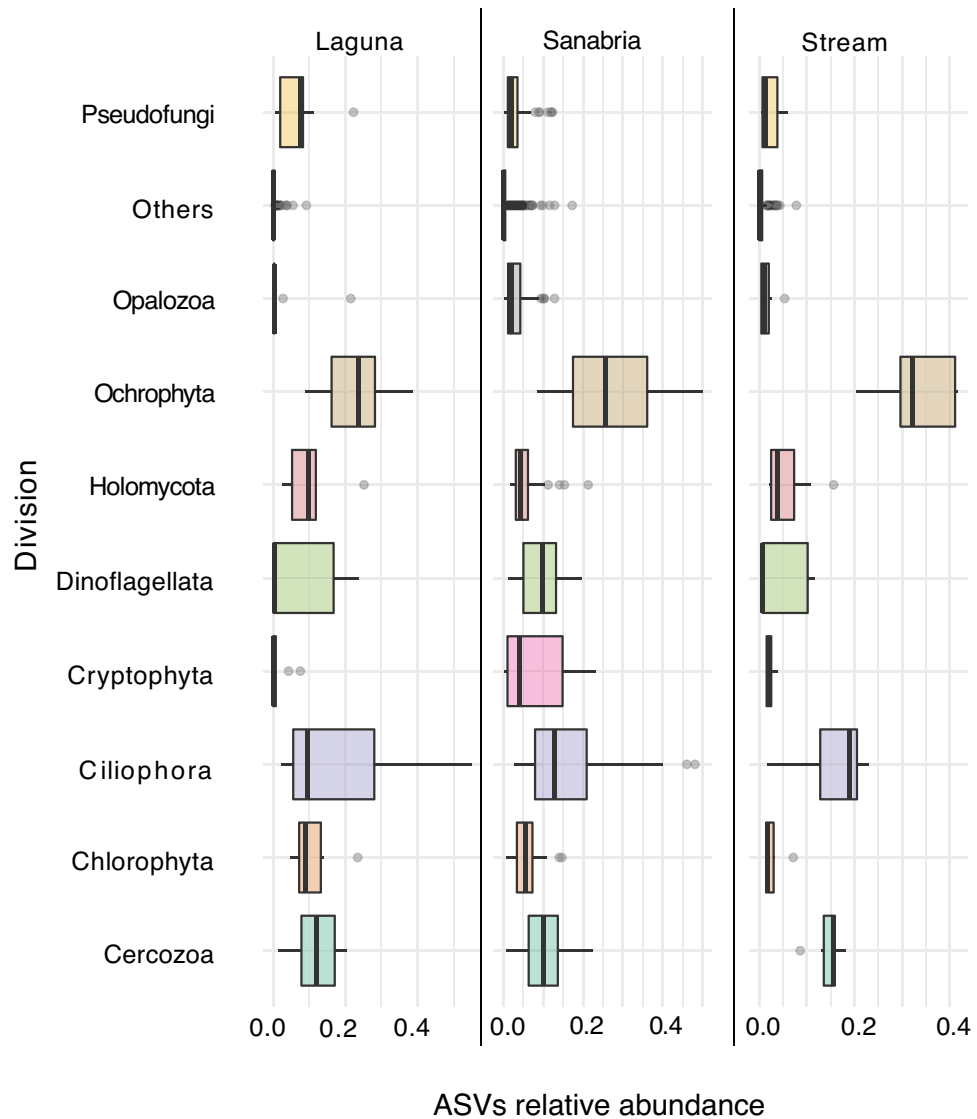


Figure 4. Distribution and relative abundance of the eukaryotic divisions across sampling sites. The taxonomy follows the system for the classification of protists proposed by Adl et al. in 2012 and implemented in the PR2 database by Guillou et al. in 2013. “Others” group together all taxa with relative abundance less than 1%. The boxes represent the interquartile range (IQR) between the first and third quartiles and the vertical line inside the box defines the median.

more than 1000 described species^{91,92}, was also the most diverse group of parasites in our study, including more than 2200 of the 5925 total parasite ASVs, distributed among more than 50 genera. Almost half of the chytrids in terms of abundance identified in our study belonged to the order Rhizophydiales, that are host-specific chytrids that infect various phytoplankton species, mostly diatoms^{93,94}. The prevalence of Rhizophydiales in the Sanabria Lake ecosystem was not surprising given that they are the most common planktonic chytrids in lacustrine ecosystems⁹⁵. A species of *Rhizophydiales* was probably the causative agent of a chytrid infection in Sanabria Lake in 2014 that diminished the population of the diatom *Tabellaria fenestrata* and controlled an algal bloom caused by eutrophication⁵⁵. The relative abundance of Perkinsea, a group of parasitic alveolates, ranged from 13 to 18% of total parasite abundance across the different sampling sites. Perkinsea were previously considered as strictly marine parasites^{96–100} until molecular environmental surveys revealed an unprecedented diversity of these organisms in freshwaters^{87,101–103}.

The parasitic community of each sampling site carried a unique taxonomic signature. The parasitic community of the tributary stream was characterised by a higher proportion of Apicomplexa (17%) and Labyrinthulomycetes (12%) in comparison to the other sampling sites. Most of the apicomplexan ASVs in the tributary stream fell into eugregarines, the most abundant apicomplexan group in environmental surveys¹⁰⁴. Parasitic Stramenopiles (Pseudofungi), a significant component of freshwater ecosystems¹⁰⁵, constituted the second most abundant group in Laguna and represented 20% of the Stramenopiles and 7% (76,070 reads) of all eukaryotes in this small pond

(Supplementary Figure 6). Within the group of parasitic Stramenopiles (Supplementary Figure 6), there was observed a higher prevalence of Oomycetes that are common fish pathogens^{106,107} in Laguna in comparison to the other sampling sites. Finally, Sanabria Lake harboured a higher relative abundance of Ichthyosporea (12%, 96,491 reads) in comparison to the other sampling sites (Stream: 2%, Laguna: 1%). The majority of the Ichthyosporea in Sanabria were associated with the marine genera *Abeoforma* (69%), *Sphaeroforma* (17%) and *Pseudoperkinsus* (10%), none of them previously identified in a freshwater environment.

To confirm the identity of the ichthyosporean ASVs in Sanabria Lake we analysed them by phylogenetic placement. We compiled a dataset that encompassed all the extant diversity of unicellular Holozoa ($n = 234$). Half of the complete 18S rDNA gene sequences used to build the reference tree belonged to uncultured environmental taxa. A total of 132 ASVs identified as Ichthyosporea by the Ribosomal Database Project (RDP) classifier were placed into the 465 branches of the reference tree (Supplementary Figure 7). Most of the queries were placed in a clade formed by the freshwater anuran parasite *Anurofeca richardsi*, the marine *Creolimax fragrantissima*, *Pseudoperkinsus tapetis* and *Sphaeroforma arctica*, and some uncultured environmental taxa (Supplementary Figure 8). The 132 ichthyosporean queries were clustered into 17 clades in the best-hit placement tree (Supplementary Figure 8). Most of the clades were associated with freshwater sequences. Clade 4, the one formed by the larger number of sequences, was assigned to the FRESHIP2 group¹⁰⁸, expanding the known molecular diversity of these environmental taxa. Clades 13, 14 and 15 were assigned to *Anurofeca richardsi* and clade 9 to *Caullerya mensii*, another freshwater parasite that infects *Daphnia pulex*¹⁰⁹. We identified two clades that were directly associated with marine Ichthyosporea, clade 6 that branched as sister to *Abeoforma whisleri* and clade 16 that branched as sister to *Sphaeroforma arctica* (Supplementary Figure 8). The genera *Abeoforma* and *Sphaeroforma* were previously considered exclusively marine and this is the first record that connects these taxa to freshwater habitats. As freshwater habitats are increasingly explored by molecular means, the number of taxa that have been previously reported as exclusively marine and later were found in freshwater surveys continues to expand^{87,110–114}.

Abundant and potentially novel freshwater microbial eukaryotes. Metabarcoding biodiversity studies have shown that a great part of the extant microbial diversity remains undocumented^{1,2}. In a metabarcoding survey, a species can be described as novel either because it is completely unknown to science or because the particular molecular marker database used in the study does not include available information on the species. In this study, we use the term ‘novelty’ to describe ASVs that are not present in our reference database as it is not possible to know for certain at this point whether or not these ASVs represent known but unsequenced species.

To check whether Sanabria Lake and its surrounding freshwater systems could be a potential sampling site to isolate new organisms, we investigated the molecular novelty by first selecting potentially novel ASVs. We used the Ribosomal Database Project (RDP) classifier¹¹⁵ to assign taxonomy to the ASVs. The RDP classifier provides for each ASV an assignment of the best matching taxa together with a bootstrap confidence score at each taxonomic rank. This score represents the level of confidence of the taxonomic assignment at each rank, from supergroup to species. Here, we define as poorly assigned, thus potentially novel, all ASVs with bootstrap confidence score value < 97 at the supergroup level. We were interested in identifying the most abundant and novel microbial eukaryotes in our study site, so we selected all ASVs with more than 1000 reads and bootstrap confidence score value lower than the aforementioned established novelty threshold.

To assign taxonomy to the queries of our dataset, we analysed them using phylogenetic placement (Fig. 5). We first constructed a comprehensive eukaryotic reference tree with 618 eukaryotic taxa that encompassed all the extant eukaryotic diversity according to the latest classification of eukaryotes⁸³. We designed the reference tree with two criteria. First, to be inclusive in order to minimise phylogenetic placement artefacts related to taxonomic sampling and second to be non-redundant in order to be smaller and thus easier to handle in the post placement analyses. The amplicon short sequences were aligned to the reference alignment and the amplicon sequences that were not aligned in the V4 region were removed as artefacts after manual inspection. We placed a total of 113 ASV V4 queries into 1233 branches of the reference tree.

Most of the ASV placements in the tree were found in the leaf nodes of Rhodophyta (Archaeplastida), Bigyra (Stramenopiles), early-branching Nucleotmycea (also known as Holomycota), and Apicomplexa (Alveolata), pinpointing these clades as parts of the tree with potential novel undescribed molecular diversity (Fig. 5). An elevated number of placements was spotted in the internal nodes of Dinophyta and the divergence between Opisthokonta and Apusomonadida (Fig. 5). Apusomonadida is a recently defined phylum with a key phylogenetic position to understand the origin of the eukaryotic cell. Apusomonads are rarely detected in environmental studies^{116–120} and can be considerably more diverse than is currently perceived¹²¹. We report previously undocumented diversity associated with the genera *Cryptomonas* and *Chilomonas* inside Cryptista, the naked filose amoebae of the genus *Vampyrella* (Endomyxa), and the frequently detected by 18S rRNA gene sequencing eukaryovorous biflagellate Aquavolon¹²². No placement was recorded inside the group of Excavata.

Conclusions

Metabarcoding analyses of the V4 hypervariable region of the 18S rRNA gene from diverse habitats in Sanabria Lake and the surrounding freshwater ecosystems uncovered a rich and diverse microeukaryotic community. One fifth of the diversity of microeukaryotes identified in Sanabria Lake are parasites, stressing the importance of parasitic taxa in the freshwater ecosystems. Our observations regarding the taxonomic composition of the microeukaryotic community overlap with microscopical data based on morphology but expand the total biodiversity recorded in the lake by adding taxa that were either insufficiently abundant to be detected by traditional methods or inconspicuous due to lack of taxonomically informative morphological characters. Tributary stream samples were significantly more species-rich than samples from Sanabria lake and Laguna. We found that

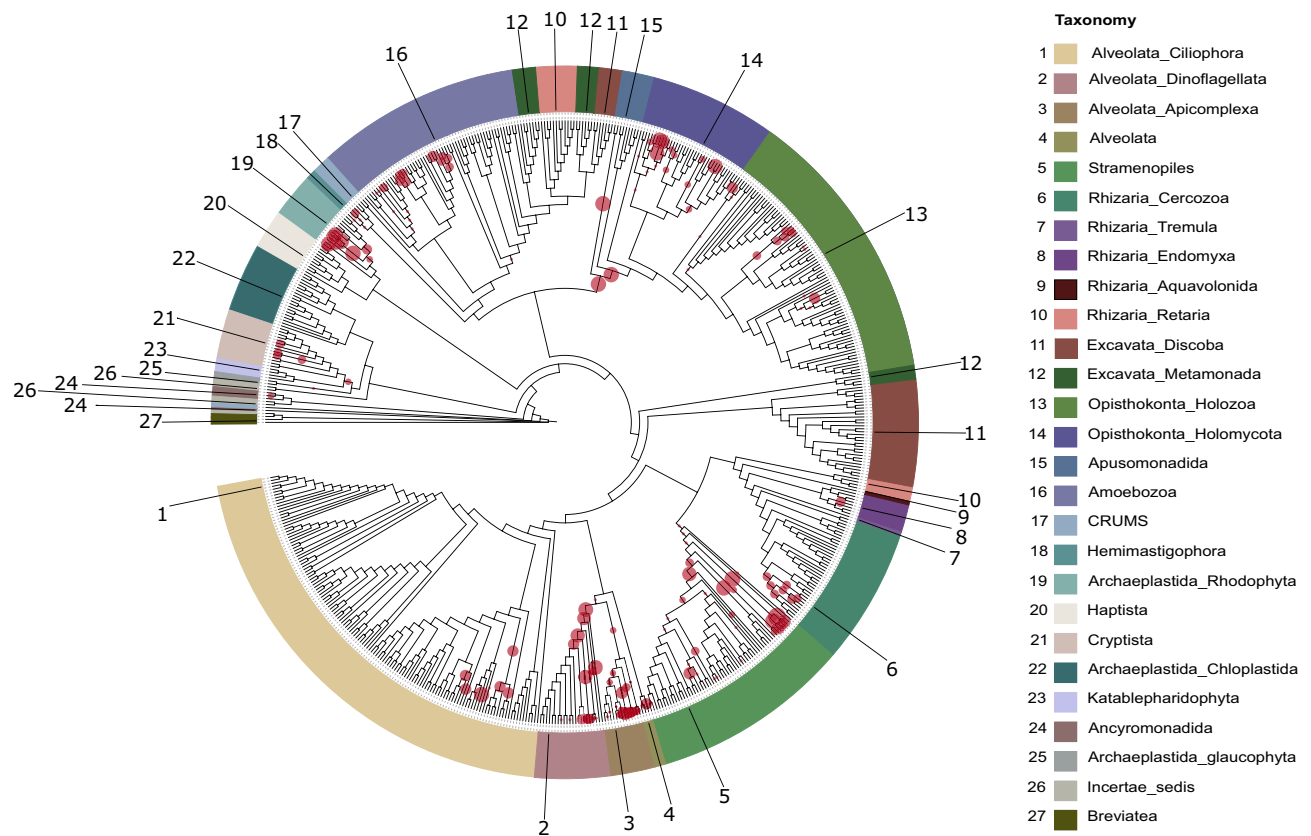


Figure 5. Novelty estimation. The tree shows the result of the phylogenetic placement of 113 ASVs into 1233 branches of a reference tree with 618 taxa. The reference taxon sampling spans all extant eukaryotic diversity as described in Adl et al. 2019. The diameter of the circles indicates the number of ASVs placed in each branch. The same ASV can be placed on multiple branches until its accumulated likelihood weight ratio reaches a value close to 1. Taxonomic groups with high number of placements in interior nodes indicate potential novel undescribed molecular diversity.

sediments harbour the greatest diversity among different habitats. We observed compositional heterogeneity among the microbial communities of Sanabria Lake and the surrounding freshwater ecosystem. Phylogenetic placement analysis showed that Sanabria Lake and the surrounding freshwater ecosystems would be good targets for future studies aiming the discovery of potential novel microeukaryotes. This is the first metabarcoding record of the diversity in Sanabria Lake. Our results expand our understanding of the microbial communities in oligomesotrophic, temperate, lacustrine ecosystems and can be used as reference for future studies in the area.

Data availability

Raw data are available in the European Nucleotide Archive (ENA) under the accession number PRJEB23911 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB23911>). Supplementary materials are available as figshare public repository under the <https://doi.org/10.6084/m9.figshare.21884934>. The code is available at https://github.com/kcmits/microeuk_Sanabria.

Received: 3 July 2022; Accepted: 18 February 2023

Published online: 22 February 2023

References

- Pawlowski, J. et al. CBOL Protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLOS Biol.* **10**, e1001419 (2012).
- del Campo, J. et al. The others: our biased perspective of eukaryotic genomes. *Trends Ecol. Evol.* **29**, 252–259 (2014).
- Handbook of the Protists* (Springer, 2017). <https://doi.org/10.1007/978-3-319-28149-0>.
- Lang, B. F., O’Kelly, C., Nerad, T., Gray, M. W. & Burger, G. The closest unicellular relatives of animals. *Curr. Biol.* **12**, 1773–1778 (2002).
- del Campo, J. et al. Ecological and evolutionary significance of novel protist lineages. *Eur. J. Protistol.* **55**, 4–11 (2016).
- Grau-Bové, X. et al. Dynamics of genomic innovation in the unicellular ancestry of animals. *Life* **6**, e26036 (2017).
- Gawryluk, R. M. R. et al. Non-photosynthetic predators are sister to red algae. *Nature* **572**, 240–243 (2019).
- Gabr, A., Grossman, A. R. & Bhattacharya, D. Paulinella, a model for understanding plastid primary endosymbiosis. *J. Phycol.* **56**, 837–843 (2020).
- Gao, Z., Karlsson, I., Geisen, S., Kowalchuk, G. & Jousset, A. Protists: Puppet masters of the rhizosphere microbiome. *Trends Plant Sci.* **24**, 165–176 (2019).

10. Caron, D. A. New accomplishments and approaches for assessing protistan diversity and ecology in natural ecosystems. *Bioscience* **59**, 287–299 (2009).
11. Gooday, A. J., Schoenle, A., Dolan, J. R. & Arndt, H. Protist diversity and function in the dark ocean: Challenging the paradigms of deep-sea ecology with special emphasis on foraminiferans and naked protists. *Eur. J. Protistol.* **75**, 125721 (2020).
12. Stoecker, D. K., Johnson, M. D., de Vargas, C. & Not, F. Acquired phototrophy in aquatic protists. *Aquat. Microb. Ecol.* **57**, 279–310 (2009).
13. Strom, S. L., Benner, R., Ziegler, S. & Dagg, M. J. Planktonic grazers are a potentially important source of marine dissolved organic carbon. *Limnol. Oceanogr.* **42**, 1364–1374 (1997).
14. Orsi, W. D. *et al.* Identifying protist consumers of photosynthetic picoeukaryotes in the surface ocean using stable isotope probing. *Environ. Microbiol.* **20**, 815–827 (2018).
15. Corno, G. & Jürgens, K. Direct and indirect effects of protist predation on population size structure of a bacterial strain with high phenotypic plasticity. *Appl. Environ. Microbiol.* **72**, 78–86 (2006).
16. Mahé, F. *et al.* Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat. Ecol. Evol.* **1**, 91 (2017).
17. Ruppert, K. M., Kline, R. J. & Rahman, M. S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* **17**, e00547 (2019).
18. Epstein, S. & López-García, P. “Missing” protists: a molecular prospective. *Biodivers. Conserv.* **17**, 261–276 (2008).
19. López-García, P., Rodríguez-Valera, F., Pedrós-Alió, C. & Moreira, D. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409**, 603–607 (2001).
20. Lovejoy, C., Massana, R. & Pedrós-Alió, C. Diversity and distribution of marine microbial eukaryotes in the Arctic Ocean and adjacent seas. *Appl. Environ. Microbiol.* **72**, 3085–3095 (2006).
21. Worden, A. Z., Cuvelier, M. L. & Bartlett, D. H. In-depth analyses of marine microbial community genomics. *Trends Microbiol.* **14**, 331–336 (2006).
22. Countway, P. D. *et al.* 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 (2007).
23. Massana, R. & Pedrós-Alió, C. Unveiling new microbial eukaryotes in the surface ocean. *Curr. Opin. Microbiol.* **11**, 213–218 (2008).
24. Alexander, E. *et al.* Microbial eukaryotes in the hypersaline anoxic L’Atalante deep-sea basin. *Environ. Microbiol.* **11**, 360–381 (2009).
25. Stoeck, T. *et al.* Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31 (2010).
26. Logares, R. *et al.* Patterns of rare and abundant marine microbial eukaryotes. *Curr. Biol.* **24**, 813–821 (2014).
27. de Vargas, C. *et al.* Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 150 (2015).
28. Fell, J. W., Scorzetti, G., Connell, L. & Craig, S. Biodiversity of micro-eukaryotes in Antarctic Dry Valley soils with <5% soil moisture. *Soil Biol. Biochem.* **38**, 3107–3119 (2006).
29. Shen, C. *et al.* Contrasting elevational diversity patterns between eukaryotic soil microbes and plants. *Ecology* **95**, 3190–3202 (2014).
30. Moon-van der Staay, S. Y. *et al.* Eukaryotic diversity in historical soil samples. *FEMS Microbiol. Ecol.* **57**, 420–428 (2006).
31. Šlapeta, J., Moreira, D. & López-García, P. The extent of protist diversity: insights from molecular ecology of freshwater eukaryotes. *Proc. R. Soc. B Biol. Sci.* **272**, 2073–2081 (2005).
32. Debroas, D. *et al.* Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available metabarcoding data. *FEMS Microbiol. Ecol.* **93**, 875 (2017).
33. Dodson, S. Predicting crustacean zooplankton species richness. *Limnol. Oceanogr.* **37**, 848–856 (1992).
34. Reche, I., Pulido-Villena, E., Baquero, R. & Casamayor, E. Does ecosystem size determine aquatic bacterial richness?. *Ecology* **86**, 1715–1722 (2005).
35. Villarino, E. *et al.* Large-scale ocean connectivity and planktonic body size. *Nat. Commun.* **9**, 142 (2018).
36. Richter, D. *et al.* Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems (2020).
37. Dias, M. S., Cornu, J.-F., Oberdorff, T., Lasso, C. A. & Tedesco, P. A. Natural fragmentation in river networks as a driver of speciation for freshwater fishes. *Ecography* **36**, 683–689 (2013).
38. Charvet, S., Vincent, W. F., Comeau, A. M. & Lovejoy, C. Pyrosequencing analysis of the protist communities in a High Arctic meromictic lake: DNA preservation and change. *Front. Microbiol.* **3**, 415 (2012).
39. Lepère, C., Domaizon, I., Hugoni, M., Vellet, A. & Debroas, D. Diversity and dynamics of active small microbial eukaryotes in the anoxic zone of a Freshwater Meromictic Lake (Pavin, France). *Front. Microbiol.* **7** (2016).
40. Boenigk, J. *et al.* Geographic distance and mountain ranges structure freshwater protist communities on a European scale. *Metabarcoding Metagenom.* **2** (2018).
41. Kammerlander, B. *et al.* High diversity of protistan plankton communities in remote high mountain lakes in the European Alps and the Himalayan mountains. *FEMS Microbiol. Ecol.* **91** (2015).
42. Filker, S., Sommaruga, R., Vila, I. & Stoeck, T. Microbial eukaryote plankton communities of high-mountain lakes from three continents exhibit strong biogeographic patterns. *Mol. Ecol.* **25**, 2286–2301 (2016).
43. Stoeck, T. *et al.* A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. *Environ. Microbiol.* **16**, 430–444 (2014).
44. de Daniel, A. C., Pedrós-Alió, C., Pearce, D. A. & Alcamí, A. Composition and Interactions among Bacterial, Microeukaryotic, and T4-like Viral Assemblages in Lakes from Both Polar Zones. *Front. Microbiol.* **7**, 337 (2016).
45. Stoof-Leichsenring, K. R., Dulias, K., Biskaborn, B. K., Pestryakova, L. A. & Herzsich, U. Lake-depth related pattern of genetic and morphological diatom diversity in boreal Lake Bolshoe Toko Eastern Siberia. *PLoS ONE* **15**, e0230284 (2020).
46. Lefranc, M., Thénot, A., Lepère, C. & Debroas, D. Genetic diversity of small eukaryotes in lakes differing by their trophic status. *Appl. Environ. Microbiol.* **71**, 5935–5942 (2005).
47. Lepère, C. *et al.* Geographic distance and ecosystem size determine the distribution of smallest protists in lacustrine ecosystems. *FEMS Microbiol. Ecol.* **85**, 85–94 (2013).
48. Vega, J. C., De Hoyos, C., Aldasoro, J. & Miguel, J. Nuevos datos morfométricos para el Lago de Sanabria. *Limnetica ISSN 0213-8409 Vol 24 Nº 1-2 2005 Ejemplar Dedic. XI Congr. Asoc. Esp. Limnol. III Congr. Ibérico Limnol. Pags 115–121* **24** (2005).
49. Margalef, R. Los organismos indicadores en la limnología. (1955).
50. De Hoyos, C. & Comín, F. The importance of inter-annual variability for management. *Hydrobiologia* **395–396**, 281–291 (1999).
51. Negro, A. I., De Hoyos, C. & Aldasoro, J. J. Diatom and desmid relationships with the environment in mountain lakes and mires of NW Spain. *Hydrobiologia* **505**, 1–13 (2003).
52. Luque, J. A. Lake sediment response to land-use and climate change during the last 1000 years in the oligotrophic Lake Sanabria (northwest of Iberian Peninsula). *Sediment. Geol.* **148**, 343–355 (2002).
53. Jambriña-Enríquez, M. *et al.* Timing of deglaciation and postglacial environmental dynamics in NW Iberia: the Sanabria Lake record. *Q. Sci. Rev.* **94**, 136–158 (2014).
54. Pahissa, J., Fernández-Enríquez, C. & De Hoyos, C. Water quality of Lake Sanabria according to phytoplankton. A comparison with historical data. *Limnetica* **10**, 527–540. <https://doi.org/10.23818/limn.34.39> (2015).

55. Llorente, A. & Seoane, S. Changes in the phytoplankton community structure in a monomictic temperate lake. *Limnetica* **39**, 469–485 (2020).
56. Vega, J. C. The Sanabria lake. The largest natural freshwater lake in Spain. *Limnetica* **8**, 49–57 (1992).
57. Edmondson, W. T. Margalef, R. 1983. *Limnología*. Ediciones Omega, S.A., Barcelona. 1010 p. *Limnol. Oceanogr.* **29**, 1349–1349 (1984).
58. Hoyos Alonso, C. de. *Limnología del lago de sanabria: variabilidad interanual del fitoplancton*. (Universidad de Salamanca, 1997).
59. Rodríguez-Rodríguez, L., Monserrat, J.-S., M.J., D.-C., Rico, M. & Valero-Garcés, B. Last deglaciation in northwestern Spain: New chronological and geomorphologic evidence from the Sanabria Region. *Geomorphology* **135**, 48–65 (2011).
60. Oterino, A. G. *Lago de sanabria, presente y futuro de un ecosistema en desequilibrio* (Antonio Guillén Oterino, 2015).
61. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
62. Guillou, L. *et al.* The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**, D597–604 (2013).
63. McMurdie, P. J. & Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
64. Weiss, S. *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 7854 (2017).
65. Xie, Y. *et al.* Environmental DNA metabarcoding reveals primary chemical contaminants in freshwater sediments from different land-use types. *Chemosphere* **172**, 201–209 (2017).
66. Pielou, E. C. The measurement of diversity in different types of biological collections. *J. Theor. Biol.* **13**, 131–144 (1966).
67. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2009). <https://doi.org/10.1007/978-0-387-98141-3>.
68. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
69. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
70. Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinform.* **11**, 538 (2010).
71. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
72. Berger, S. A. & Stamatakis, A. Aligning short reads to reference alignments and trees. *Bioinformatics* **27**, 2068–2075 (2011).
73. Letunic, I. & Bork, P. Interactive Tree Of Life v2: Online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**, W475–478 (2011).
74. Çako, V., Baci, S. & Shena, M. Water turbidity as one of the trophic state indices in Butrinti Lake. *J. Water Resour. Prot.* **05**, 1144–1148 (2013).
75. Chanudet, V. & Filella, M. Submicron organic matter in a peri-alpine, ultra-oligotrophic lake. *Org. Geochem.* **38**, 2545 (2007).
76. Poikane, S. *et al.* Defining ecologically relevant water quality targets for lakes in Europe. *J. Appl. Ecol.* **51**, 592–602 (2014).
77. Ministerio de Agricultura, Alimentación y Medio Ambiente. *Real Decreto 817/2015, de 11 de septiembre, por el que se establecen los criterios de seguimiento y evaluación del estado de las aguas superficiales y las normas de calidad ambiental*. vol. BOE-A-2015-9806 80582-80677 (2015).
78. Dunthorn, M., Klier, J., Bunge, J. & Stoeck, T. Comparing the hyper-variable V4 and V9 regions of the small subunit rDNA for assessment of ciliate environmental diversity. *J. Eukaryot. Microbiol.* <https://doi.org/10.1111/j.1550-7408.2011.00602.x> (2012).
79. Williams, P. *et al.* Comparative biodiversity of rivers, streams, ditches and ponds in an agricultural landscape in Southern England. *Biol. Conserv.* **115**, 329–341 (2004).
80. Eckert, E. M. *et al.* Different substrates within a lake harbour connected but specialised microbial communities. *Hydrobiologia* **847**, 1689–1704 (2020).
81. Piredda, R. *et al.* Diatom diversity through HTS-metabarcoding in coastal European seas. *Sci. Rep.* **8**, 18059 (2018).
82. Adl, S. M. *et al.* The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59**, 429–514 (2012).
83. Adl, S. M. *et al.* Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.* **66**, 4–119 (2019).
84. Nicholls, K. H. & Wujek, D. E. 12—Chrysophycean algae. in *Freshwater Algae of North America* (eds. Wehr, J. D. & Sheath, R. G.) 471–509 (Academic Press, 2003). <https://doi.org/10.1016/B978-012741550-5/50013-1>.
85. Cavalier-Smith, T., Chao, E. E. & Lewis, R. Multiple origins of Heliozoa from flagellate ancestors: New cryptist subphylum Corbihelia, superclass Corbistoma, and monophyly of Haptista, Cryptista, Hacrobia and Chromista. *Mol. Phylogenet. Evol.* **93**, 331–362 (2015).
86. Okamoto, N. & Inouye, I. The Katablepharids are a distant sister group of the cryptophyta: A proposal for Katablepharidophyta Divisio nova/kathablepharida phylum novum based on SSU rDNA and beta-tubulin phylogeny. *Protist* **156**, 163–179 (2005).
87. Bråte, J. *et al.* Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *ISME J.* **4**, 1144–1153 (2010).
88. Frenken, T. *et al.* Integrating chytrid fungal parasites into plankton ecology: research gaps and needs. *Environ. Microbiol.* **19**, 3802–3822 (2017).
89. Lefèvre, E. *et al.* Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: evidence from a molecular diversity study in a deep meromictic lake. *Environ. Microbiol.* **9**, 61–71 (2007).
90. Sime-Ngando, T., Lefevre, E. & Gleason, F. Hidden diversity among aquatic heterotrophic flagellates: Ecological potentials of zoospore fungi. *Hydrobiologia* **659** (2011).
91. Powell, M. J. Looking at mycology with a Janus face: A glimpse at chytridiomycetes active in the environment. *Mycologia* **85**, 1–20 (1993).
92. Shearer, C. A. *et al.* Fungal biodiversity in aquatic habitats. *Biodivers. Conserv.* **16**, 49–67 (2007).
93. Jobard, M., Rasconi, S. & Sime-Ngando, T. Diversity and functions of microscopic fungi: A missing component in pelagic food webs. *Aquat. Sci.* **72** (2010).
94. Rasconi, S., Jobard, M. & Sime-Ngando, T. Parasitic fungi of phytoplankton: ecological roles and implications for microbial food webs. *Aquat. Microb. Ecol.* **62**, 123–137 (2011).
95. Monchy, S. *et al.* Exploring and quantifying fungal diversity in freshwater lake ecosystems using rDNA cloning/sequencing and SSU tag pyrosequencing. *Environ. Microbiol.* **13**, 1433–1453 (2011).
96. Norén, F., Moestrup, Ø. & Rehnstam-Holm, A.-S. Parvilucifera infectans norén et moestrup gen. et sp. nov. (perkinsozoa phylum nov.): a parasitic flagellate capable of killing toxic microalgae. *Eur. J. Protistol.* **35**, 233–254 (1999).
97. Erard-Le Denn, E., Chrétiennot-Dinet, M.-J. & Probert, I. First report of parasitism on the toxic dinoflagellate *Alexandrium minutum* Halim. *Estuar. Coast. Shelf Sci.* **50**, 109–113 (2000).
98. Villalba, A., Reece, K. S., Ordás, M. C., Casas, S. M. & Figueras Huerta, A. *Perkinsosis in molluscs: A review*. <https://doi.org/10.1051/alr:2004050> (2004).
99. Figueroa, R. I., Garcés, E., Massana, R. & Camp, J. Description, Host-specificity, and Strain Selectivity of the Dinoflagellate Parasite Parvilucifera sinerae sp.nov. (Perkinsozoa). (2008) <https://doi.org/10.1016/j.protis.2008.05.003>.

100. Leander, B. S. & Hoppenrath, M. Ultrastructure of a novel tube-forming, intracellular parasite of dinoflagellates: Parvilucifera prorocentri sp. nov. (Alveolata, Myzozoa). *Eur. J. Protistol.* **44**, 55–70 (2008).
101. Richards, T. A., Vepritskiy, A. A., Gouliamova, D. E. & Nierzwicki-Bauer, S. A. The molecular diversity of freshwater picoeukaryotes from an oligotrophic lake reveals diverse, distinctive and globally dispersed lineages. *Environ. Microbiol.* **7**, 1413–1425 (2005).
102. Lefèvre, E., Roussel, B., Amblard, C. & Sime-Ngando, T. The molecular diversity of freshwater picoeukaryotes reveals high occurrence of putative parasitoids in the plankton. *PLoS ONE* **3**, e2324 (2008).
103. Lepère, C., Domaizon, I. & Debroas, D. Unexpected importance of potential parasites in the composition of the freshwater small-eukaryote community. *Appl. Environ. Microbiol.* **74**, 2940–2949 (2008).
104. del Campo, J. *et al.* Validation of a universal set of primers to study animal-associated microeukaryotic communities. *Environ. Microbiol.* **21**, 3855–3861 (2019).
105. Cooper, J. A., Pillinger, J. M. & Ridge, I. Barley straw inhibits growth of some aquatic saprolegniaceous fungi. *Aquaculture* **156**, 157–163 (1997).
106. van West, P. *Saprolegnia parasitica*, an oomycete pathogen with a fishy appetite: new challenges for an old problem. *Mycologist* **20**, 99–104 (2006).
107. Phillips, A. J., Anderson, V. L., Robertson, E. J., Secombes, C. J. & van West, P. New insights into animal pathogenic oomycetes. *Trends Microbiol.* **16**, 13–19 (2008).
108. del Campo, J. & Ruiz-Trillo, I. Environmental survey meta-analysis reveals hidden diversity among unicellular opisthokonts. *Mol. Biol. Evol.* **30**, 802–805 (2013).
109. Lu, Y. *et al.* Revisiting the phylogenetic position of *Caullerya mesnili* (Ichthyosporea), a common Daphnia parasite, based on 22 protein-coding genes. *Mol. Phylogenet. Evol.* **151**, 106891 (2020).
110. Simon, M. *et al.* Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. *ISME J.* **9**, 1941–1953 (2015).
111. Richards, T. A. & Bass, D. Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. *Curr. Opin. Microbiol.* **8**, 240–252 (2005).
112. Annenkova, N. V., Giner, C. R. & Logares, R. Tracing the origin of planktonic protists in an ancient lake. *Microorganisms* **8**, 543 (2020).
113. Yi, Z. *et al.* High-throughput sequencing of microbial eukaryotes in Lake Baikal reveals ecologically differentiated communities and novel evolutionary radiations. *FEMS Microbiol. Ecol.* **93**, 78458 (2017).
114. Mukherjee, I. *et al.* Widespread dominance of kinetoplastids and unexpected presence of diplomonids in deep freshwater lakes. *Front. Microbiol.* **10** (2019).
115. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
116. López-García, P., Philippe, H., Gail, F. & Moreira, D. Autochthonous eukaryotic diversity in hydrothermal sediment and experimental microcolonizers at the Mid-Atlantic Ridge. *Proc. Natl. Acad. Sci. USA* **100**, 697–702 (2003).
117. Not, F. *et al.* Protistan assemblages across the Indian Ocean, with a specific emphasis on the picoeukaryotes. *Deep Sea Res. Part Oceanogr. Res. Pap.* **55**, 1456–1473 (2008).
118. Takishita, K., Yubuki, N., Kakizoe, N., Inagaki, Y. & Maruyama, T. Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from raw sediment samples and two enrichment cultures. *Extrem. Life Extreme Cond.* **11**, 563–576 (2007).
119. Orsi, W., Song, Y. C., Hallam, S. & Edgcomb, V. Effect of oxygen minimum zone formation on communities of marine protists. *ISME J.* **6**, 1586–1601 (2012).
120. Lesaulnier, C. *et al.* Elevated atmospheric CO₂ affects soil microbial diversity associated with trembling aspen. *Environ. Microbiol.* **10**, 926–941 (2008).
121. Torruella, G., Moreira, D. & López-García, P. Phylogenetic and ecological diversity of apusomonads, a lineage of deep-branching eukaryotes. *Environ. Microbiol. Rep.* **9**, 113–119 (2017).
122. Bass, D. *et al.* Rhizarian ‘novel clade 10’ revealed as abundant and diverse planktonic and terrestrial flagellates, including aquavolon n. gen. *J. Eukaryot. Microbiol.* **65**, 828–842 (2018).

Acknowledgements

This work was supported by Grants BFU2017-90114-P from Ministerio de Economía y Competitividad (MINECO), Agencia Estatal de Investigación (AEI), and Fondo Europeo de Desarrollo Regional (FEDER) and PID2020-120609GB-I00 by MCIN/AEI/10.13039/501100011033 and “ERDF A way of making Europe” by the “European Union” to IR-T. It has also received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. H2020-MSCA-ITN-2015-675752 (SINGEK Project, <http://www.singek.eu/>). This project has also received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant Agreement No. 949745) and the support of a fellowship from “la Caixa” Foundation (ID 100010434) with fellowship code LCF/BQ/PI19/11690008. We thank Tomás Pérez Tonda, María José Orozco López, Laura Carquijero López and David Salvador Velasco for providing help during sampling.

Author contributions

A.S.A., D.L.E., M.A., and A.G. designed the sampling strategy, collected the data and revised the manuscript. K.M. and A.S.A. performed DNA extractions. K.M. designed and conducted the data analyses, interpreted the results, designed the figures and wrote the draft and the revised manuscript. D.J.R. interpreted the results and revised the manuscript. I.R.-T. conceived and designed the study, supervised the work and revised the manuscript. All authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-30228-4>.

Correspondence and requests for materials should be addressed to K.M.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

© The Author(s) 2023