environmental microbiology

Environmental Microbiology (2021) **00**(00), 00–00 doi:10.1111/1462-2920.15838

Freshwater protists: unveiling the unexplored in a large floodplain system

Sebastian Metz[1,2*] **Paula Huber**[3,4] **Victoria Accattatis**[3] **Adriana Lopes dos Santos**[5] **Estelle Bigeard**[6] **Fernando Unrein**[2] **Aurélie Chambouvet**[1] **Fabrice Not**[6] **Enrique Lara**[7] **and Melina Devercelli**[3*]

[1]*Univ Brest, CNRS, IRD, Ifremer, LEMAR, Plouzané, F-29280, France.*

[2]*Instituto Tecnológico de Chascomús (INTECH), UNSAM-CONICET, Chascomús, Buenos Aires, Argentina.*

[3]*Departamento de Hidrobiologia, Universidade Federal de São Carlos (UFSCar), Rodovia Washington Luiz, São Carlos, São Paulo, 13565-905, Brazil.*

[4]*Instituto Nacional de Limnología (INALI), Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional del Litoral, Ciudad Universitaria, Santa Fe, Argentina.*

[5]*Asian School of the Environment, Nanyang Technological University, Singapore, Singapore.*

[6]*Sorbonne Université, CNRS, Laboratoire Adaptation et Diversité en Milieu Marin UMR7144, Station Biologique de Roscoff, Roscoff, 29680, France.*

[7]*Real Jardín Botánico de Madrid, CSIC, Madrid, 28014, Spain.*

Summary

Protists play a fundamental role in all ecosystems, but we are still far from estimating the total diversity of many lineages, in particular in highly diverse environments, such as freshwater. Here, we survey the protist diversity of the Paraná River using metabarcoding, and we applied an approach that includes sequence similarity and phylogeny to evaluate the degree of genetic novelty of the protists' communities against the sequences described in the reference database PR^2. We observed that ~28% of the amplicon sequence variants were classified as novel according to their similarity with sequences from the reference database; most of them were related to heterotrophic groups traditionally overlooked in freshwater systems. This lack of knowledge extended to those groups within the green algae (Archaeplastida) that are well documented such as Mamiellophyceae, and also to the less studied Pedinophyceae, for which we found sequences representing novel deep-branching clusters. Among the groups with potential novel protists, Bicosoecida (Stramenopiles) were the best represented, followed by *Codosiga* (Opisthokonta), and the Perkinsea (Alveolata). This illustrates the lack of knowledge on freshwater planktonic protists and also the need for isolation and/or cultivation of new organisms to better understand their role in ecosystem functioning.

Introduction

Most of the eukaryotic diversity is found among single-celled organisms. Protists represent most major lineages of the tree of life with astonishing high numbers of species. They have developed different trophic strategies ranging from autotrophs to mixotrophs and obligate heterotrophs. Autotrophs contribute to a large fraction of primary production in aquatic systems (Field *et al*., 1998; Falkowski *et al*., 2003), whereas heterotrophs and mixotrophs contribute in regulating the structure of bacterial, fungi, and other small eukaryotes populations (Strom, 2008; Geisen *et al*., 2018; Gao *et al*., 2019; Bock *et al*., 2020). Protists are crucial for the global carbon cycle, respiration, and drive all major, global scale biogeochemical processes (Worden *et al*., 2015; Keeling and del Campo, 2017). Furthermore, they are involved in complex ecological interactions as mutualists or parasites of plants, animals and other protists (Bjorbækmo *et al*., 2020). Despite their central role in ecosystem functioning, we still have a limited understanding of the diversity of many protist lineages. Thus, increasing our understanding of small protists represents a principal challenge in microbial ecology.

Environmental molecular surveys have revealed a large diversity among microeukaryotes, as well the existence of several phylogenetically consistent groups that had never been isolated before (e.g. Moon-van der Staay

Received 22 July, 2021; revised 23 October, 2021; accepted 31 October, 2021. *For correspondence. Tel. 33 601 000798; Fax: 54 342 4511648; E-mail sebastian.metz@univ-brest.fr. Tel. 54 932 4233076; Fax: 54 342 4511648; E-mail meldevercelli@gmail.com

et al., 2001; Corlett, 2017; Debroas *et al*., 2017; Pedrós-Alió *et al*., 2018). Some of these groups form new deeply branching clades, which are likely to represent evolutionary intermediates between the established supergroups (Burki *et al*., 2020). The seminal work of Massana *et al*. (2004) disclosed a wealth of deep branching heterotrophic Stramenopiles in marine environments, whose ecosystem functions were later characterized (Seeleuthner *et al*., 2018). Another early work based almost exclusively on sequence analysis revealed a group of organisms basal to all fungi (Lara *et al*., 2010) that shed light on the origins of fungi and microsporidia (Torruella *et al*., 2018). Picozoa represents a novel lineage of heterotrophic picoeukaryotes, that surprisingly is positioned next to the Archaeplastida supergroup (Not *et al*., 2007; Seenivasan *et al*., 2013; Moreira and López-García, 2014). Even within well-known groups such as the Trebouxiophyceae (containing *Chlorella* and most lichen photosymbionts), a survey of the diversity showed unknown deep-branching clades in the marine plankton (Metz *et al*., 2019b). These new clades eventually found their way into curated repositories PR^2 (Guillou *et al.,* 2013), facilitating the annotation of environmental sequences from further studies. It becomes then feasible to investigate directly the environments with the greatest potential for new organisms discovery (del Campo *et al*., 2018).

Environmental DNA-based surveys still bring evidence for the existence of undiscovered diversity in protists. Recent works reveal the existence of novel deep-branching groups (Lara *et al*., 2017; Arroyo *et al*., 2018; Annenkova *et al*., 2020). It becomes then interesting to quantify the amount of novel diversity brought by environmental molecular surveys, especially when previously underexplored environments are prospected.

Several approaches are used to illustrate the level of genetic novelty present in a given environment. Most often, the sequences [operational taxonomic units (OTUs) or amplicon sequence variants (ASVs)] are aligned with their closest matches in public databases, and percentages of similarity are computed (Massana *et al*., 2011; del Campo and Massana, 2011). Another approach consists in placing the ASVs into reference phylogenetic trees, thus allowing the computation of phylogenetic distances (e.g. Matsen *et al*., 2010; Dunthorn *et al*., 2014; Mahé *et al*., 2017; Metz *et al*., 2019b). Gene similarity network analysis combines the similarity between the query ASVs and with reference sequences (Lynch *et al*., 2012; Filker *et al*., 2015; Filker *et al*., 2016; Arroyo *et al*., 2020). Arroyo *et al*. (2020) use different network metrics to compare the divergence between both sequences, providing a direct overview of the groups of ASVs that probably represent novel taxonomic groups. Most of these approaches are based on the quality of the taxonomic annotation of the public databases, which include the sum of all published knowledge on molecular diversity of eukaryotes. Still, they do not allow distinguishing between sequences from known organisms (either cultured or isolated from the environment and taxonomically identified) and other environmental sequences. This information is crucial, because functional characterization of environmental sequences can only be achieved based on their similarity with identified organisms. Nowadays, the PR^2 database (Guillou *et al*., 2013) is one of the most commonly used for eukaryotes taxonomic annotations, because it integrates valuable information from different sources, such as GenBank and EukRef (del Campo *et al*., 2018), and is manually curated by experts, providing a highly accurate source for reference sequences selection of culture or environmental taxonomic groups.

Here, we propose an approach to quantify the amount of novel genetic diversity that compares each sequence with (i) cultured/isolated organisms and (ii) environmental sequences from other studies, using the PR^2 database as a reference. We applied this approach to quantify the novelty found in the communities of planktonic microeukaryotes living in the understudied middle course of the Paraná fluvial system.

The Paraná River is one of the world's mega rivers that runs along almost 3900 km, crossing from tropical to warm-temperate climate zones (Roberto *et al*., 2009). At its middle stretch, the main channel is fringed by a large floodplain that hosts a network of lotic and lentic environments interconnected by hydrological fluctuations (Supporting Information Fig. S1). Few studies on molecular protist diversity have been undertaken in Neotropical floodplain rivers (Lentendu *et al*., 2018; Machado *et al*., 2019) and only one study was carried out in the Paraná River, specifically in the middle stretch focussed on Opisthokonta, the supergroup formed by animals, fungi and the ancestors of the multicellular organisms (Arroyo *et al*., 2018).

In this work, we performed a metabarcoding survey with the goal of unveiling the diversity and genetic novelty of the protists at the Paraná fluvial system and extending our current knowledge of freshwater environments. We used Illumina technology for sequencing of the V4 region of the 18S rRNA gene, sequences similarity analysis, and phylogeny. Because of the complexity of the systems, with lotic and lentic environments interconnected by hydrological fluctuations, we hypothesize that it harbours a great diversity of protists, and we expect finding high level of genetic novelty. This applies particularly to heterotrophic protists, consumers and parasites that were traditionally overlooked in these environments. Because neotropical floodplains have scarcely been studied with metabarcoding approaches, we predict that a large

amount of diversity will also be far related from environmental sequences derived from other approaches.

Results

High-throughput sequences processing and community composition

In total, we obtained 1 573 449 reads grouped in 5666 amplicon sequence variants (ASVs). After excluding ASVs affiliated to Metazoa, Streptophyta, chimeras and those with less than 10 reads distributed in less than three samples, the final table comprised 3166 ASVs and 1 227 460 reads distributed in 10 samples from lentic and lotic environments from the Paraná River floodplain system (Supporting Information Fig. S1).

The ASVs were affiliated to nine supergroups according to the definition of Adl *et al*. (2019) with variable diversity and abundance (Fig. 1 and Table S1). The richest supergroups were Stramenopiles (28.04% of the contribution to total ASVs), Alveolata (22.36% ASVs), and Cryptista (15.19% ASVs, represented mostly by the class Cryptophyceae). Within Stramenopiles, the most diverse groups were Bicosoecida (10.92% ASVs) and Chrysophyceae (8.46% of ASVs). Among Alveolata, the most represented taxonomic groups were Perkinsea and the ciliate class Spirotrichea (∼5.24% and 4.89% ASVs, respectively). Archaeplastida (16.96% ASVs) and Opisthokonta (12.79% ASVs) contributed less but had a significant number of ASVs. The remaining were related to Rhizaria (2.96% ASVs), Haptista (0.82% ASVs), Amoebozoa (0.6% ASVs), and Apusomonadida (0.25% ASVs).

The hydrological period studied corresponded to drought (years 2013 and 2014), flood (year 2015), and extreme flood (year 2016) (Supporting Information Table S2). The Non-metric multidimensional scaling (NMDS) showed a separation between samples from different years, grouping together the samples from 2016 indicating a high similarity among protist community structure from extreme flood hydrological phase (Supporting Information Fig. S2). The alpha diversity analysis showed a mean Shannon index (H) equal to 4.33 ± 0.52, with a max of 5.03 corresponding to extreme flood period (year 2016) and a min of 3.49 during drought period (year 2014). The mean beta diversity using Bray–Curtis distance was 0.78 ± 0.17 (Supporting Information Table S4), indicating a high heterogeneity among protist communities at each environment and hydrological period. The lowest distance (0.26) was shown among protist diversity from extreme flood period.

The ecological function of the ASVs was manually assigned to consumer, parasitic, phototrophic or saprobic following Singer *et al*. (2021). The consumers represented ∼53% of the total ASVs, followed by phototrophic (∼36% of the ASVs), parasitic and saprobic, which

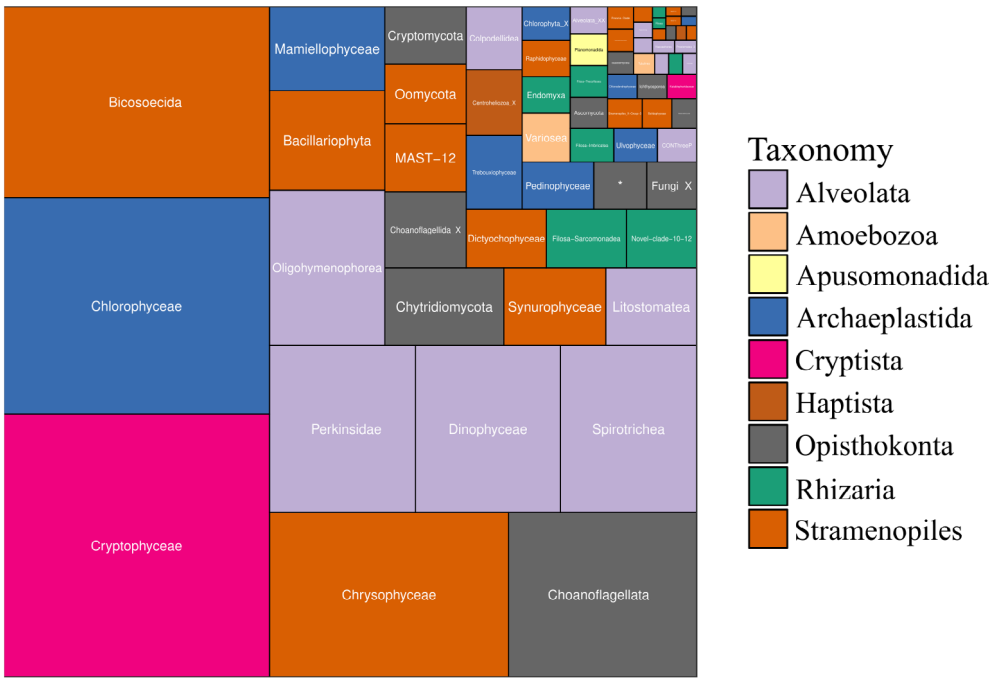

Fig. 1. Taxonomic composition of freshwater protists from the Paraná fluvial system. Relative richness (number of ASVs) of each taxonomic group were classified among the nine recognized supergroups identified by colour. The * corresponds to those groups of ASVs whose taxonomic affiliation could not be reached at class level. [Color figure can be viewed at wileyonlinelibrary.com]

represented ~9% and ~1% of the ASVs, respectively. The rest (~1%) could not be assigned to any of these categories (Supporting Information Table S3).

Novel diversity analysis

The genetic similarity between the ASVs and the PR^2 version 4.12 (Guillou *et al*., 2013), which contains environmental and culture/isolated derived sequences, varied notably among supergroups (Fig. 2). Cryptista, Alveolata, and Archaeplastida presented a unimodal distribution, with a peak close to 100% of similarity. Alveolata presented a longer tail as more ASVs with low similarity in comparison with the other two supergroups. Stramenopiles and Rhizaria had a bimodal distribution, with peaks at 90% and 99% of similarity, respectively. Opisthokonta peaked at 79% of similarity, as we obtained a high number of ASVs with poor matches in the PR^2 database.

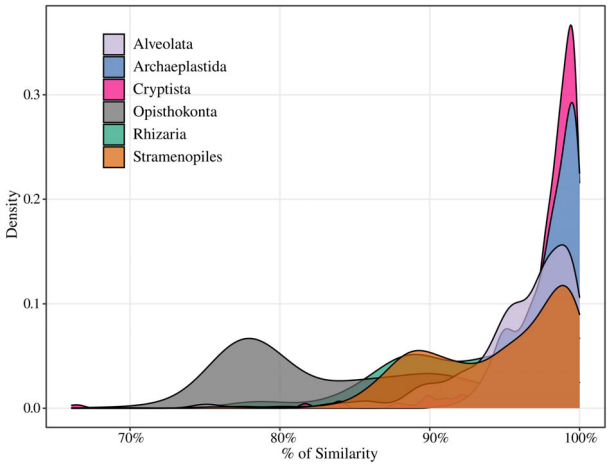

Fig. 2. Distribution of the similarity between the ASVs from each of the principal supergroups found in the Paraná fluvial system with the whole PR^2 database (both cultures and environmental sequences). [Color figure can be viewed at wileyonlinelibrary.com]

To compare the degree of novelty between the supergroups, after assigning each ASV to the closest culture/isolated match (CIM) and the closest environmental match (CEM), we classified the ASVs according to their position in the biplot analysis, showing distinct novelty patterns (Fig. 3). ASVs were assigned to quadrant I when both of their genetic similarity with CIM and CEM were above the average of the CIM and CEM of the complete database (i.e. taking into account the similarity from all the ASVs with the database), quadrant II when only CIM was above the average, quadrant III when ASVs were below average in both CIM and CEM, and quadrant IV when only CEM was below the average. Quadrant III displays, therefore, the ASVs related to potential novel protists (i.e. ASVs without closest reference sequences in the PR^2 database). In the global comparison, the average of the CIM and CEM was 91.2% and 94.2%, respectively (Fig. 3). Among the most represented supergroups, Opisthokonta showed a highest percentage of ASVs in the quadrant III (77% of ASVs, $n = 312$), following by Stramenopiles (36% of ASVs, $n = 171$) and Alveolata (29% of the ASVs, $n = 209$). Cryptista and Archaeplastida showed 5% and 0.5% of the ASVs in the quadrant III, respectively (Fig. 3).

Following the global analysis of novelty, we calculated the average of the CIM and CEM values for each supergroup (i.e. only taking into account the CIM and CEM assigned to each supergroup, Fig. 4). With these new values, the percentage of novel ASVs in quadrant III

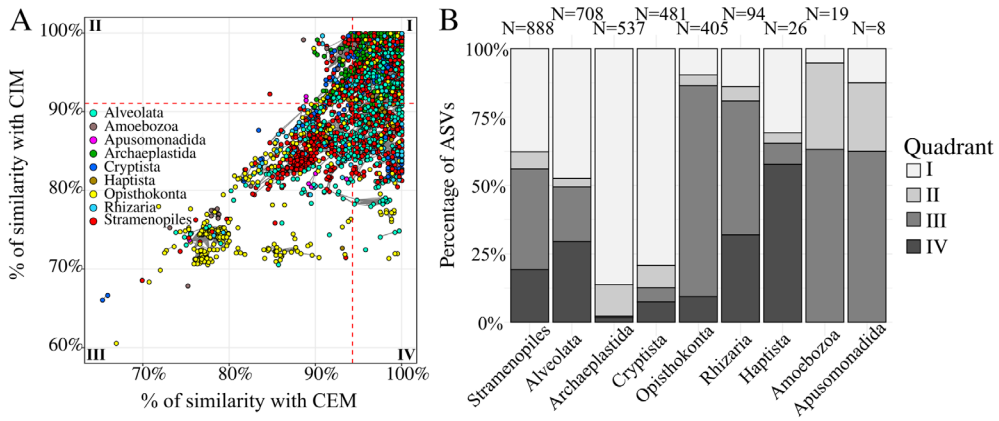

Fig. 3. Novel diversity analysis for each of the most abundant and richest supergroups of freshwater protists in the Paraná fluvial system.
A. Similarity analysis. The nodes represent the ASVs and the colours represent the affiliation. The ASVs with more than 95% of similarity are connected by edges (grey lines). The dashed red line is the average of the similarities of the ASVs with the CIMs and CEMs.
B. The bar plot represents the number of ASVs to each quadrant of the biplot per taxonomic group. ASVs from quadrant I and II were classified as 'non-novel ASVs', quadrant III as 'novel ASVs', and quadrant IV as 'environmental ASV'. [Color figure can be viewed at wileyonlinelibrary.com]

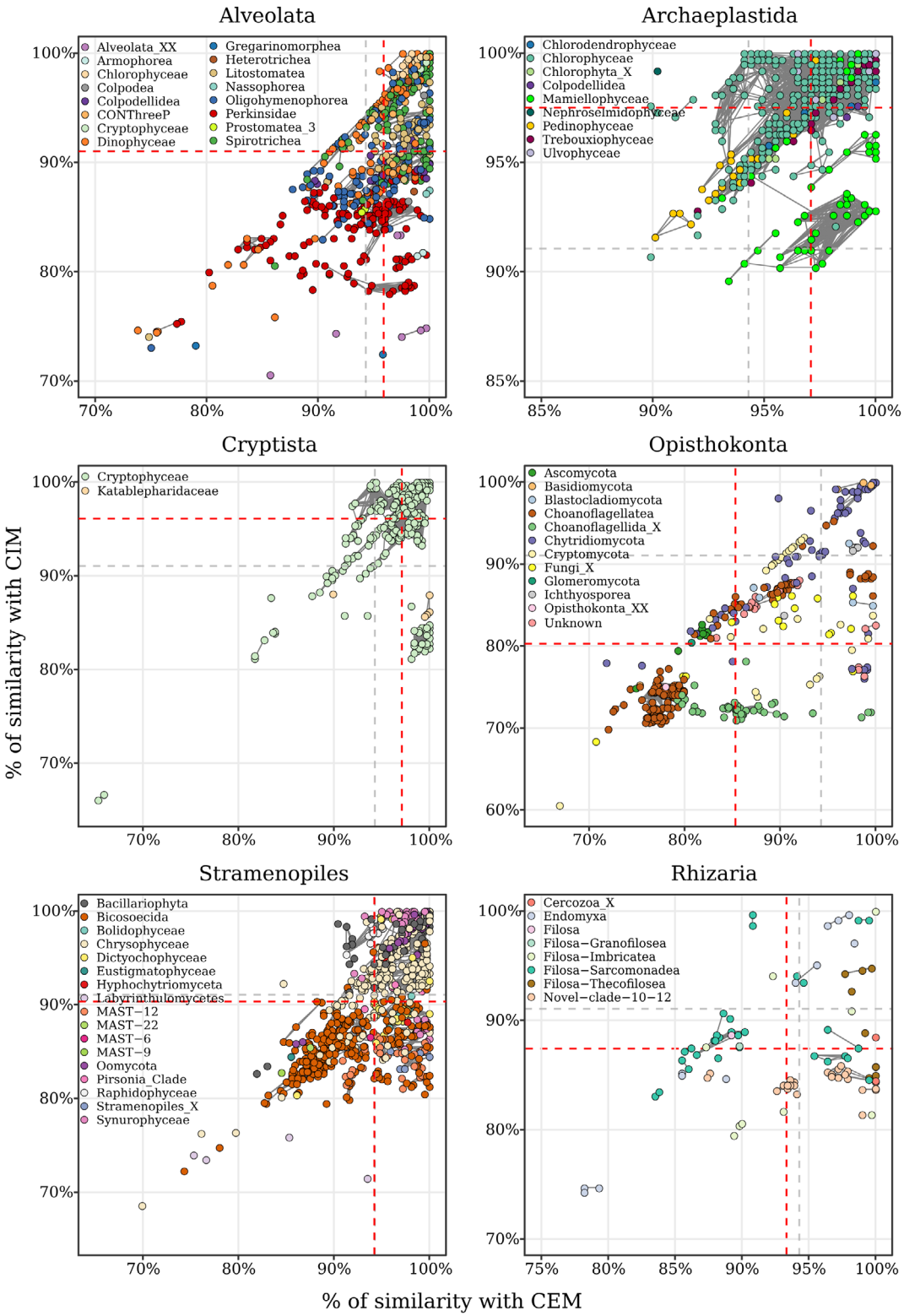

Fig. 4. Similarity analysis for the main supergroups. The nodes represent the ASVs and the colours represent the affiliation. The ASVs with more than 95% of similarity are connected by edges (grey lines). The dashed red line is the average of the similarities of the ASVs with the CIMs and CEMs for the specific supergroup analysed. The dashed grey line is the average of the similarities of the ASVs with the CIMs and CEMs in the entire database. [Color figure can be viewed at wileyonlinelibrary.com]

related to Opisthokonta decreased considerably, to 46.4% (188 ASVs), and the number of novel ASVs for Cryptista and Archaeplastida increased to 15.7% (76 ASVs) and 28.1% (152 ASVs), respectively. Among Opisthokonta, the novel ASVs were affiliated mostly to Choanoflagellata (160 ASVs). Most of the Stramenopiles' novel ASVs were classified as Bicosoecida (266 ASVs of 320 ASVs) and Chrysophyceae (28 ASVs). Perkinsea (105 ASVs of 209 ASVs) and Ciliophora (56 ASVs) were the most novel groups among Alveolata. For Archaeplastida, the novel ASVs were affiliated to Chlorophyceae (97 ASVs), Mamiellophyceae (26 ASVs), and Pedinophyceae (21 ASVs).

Despite that the similarity cut-off threshold for species-level distinction in protists is a matter of debate, previous results showed that 98% of similarity is a reasonable value for most of them (Caron *et al*., 2009). Here, we were more conservatives and ASVs with 95% of similarity between them were grouped in different clusters to capture closely related taxa (Fig. 4). Bicosoecida presented 188 clusters of ASVs and Chrysophyceae presented 73 clusters. Among Alveolata, the biggest cluster corresponded to Spirotrichea (77 ASVs), whereas Perkinsea and Oligohymenophorea showed 47 and 48 clusters, respectively. The Archaeplastida supergroup showed highly related ASVs, with only 52 clusters. Similar results were observed in Cryptista, with which only 35 clusters were obtained. On the other hand, Opisthokonta presented 167 clusters, and the most diverse was related to Choanoflagellata (formed by 57 different ASVs), which also presented a low similarity with the CIM and CEM (Fig. 4).

Phylogenetic affiliation of novel ASVs

The phylogenetic analysis of the clusters of novel ASVs related to Choanoflagellata, Perkinsea, and Bicosoecida was performed by phylogenetic placement of the ASVs sequences in a reference tree (Fig. 5). To classify the Choanoflagellata novel ASVs (160 ASVs), we used the reference sequences from Arroyo *et al*. (2018) (Supplementary dataset 4). The model selected by IQ-Tree (Nguyen *et al.,* 2015; Kalyaanamoorthy *et al*., 2017; Minh *et al*., 2020) for the phylogeny reconstruction was TN + F + R6. Most of the ASVs were placed in two big clusters related to Craspedida, more specifically closely related to *Codosiga hollandica* (KT757430) and *Codosiga* sp. M5 (JF706239) reference sequences. One of these clusters is formed by 16 different groups of ASVs with 95% of similarity and the other one by 4 (Fig. 5A and Fig. S3). These results highlight a high genetic variability in the group with a high potential to discover new species and extend the actual known diversity of genus *Codosiga*.

For Bicosoecida, the 269 novel ASVs sequences were placed in a reference tree inferred by 141 sequences. The best model for this phylogenetic reconstruction was TNe + R4. The reference sequences were selected from the bibliography to get a good representation of the different clades (del Campo and Massana, 2011; Schoenle *et al*., 2020). The ASVs were distributed along all the subgroups of this taxonomic group (Fig. 5B and Fig. S4). The phylogeny reveals that most of the ASVs were Pseudodendromonadales. Furthermore, the phylogenetic placement revealed groups of ASVs that branched at the base of common marine clades, such as Cacitellaceae (ASV_352 and ASV_563), Symbiomonadaceae (ASVs from the cluster 47, 106, 52 and ASV_817), and Cafeteriaceae (ASVs from the cluster 140, 42 and ASV_584) (Fig. 5B and Fig. S4). The long branch at the base of these clades and the low similarity of the ASVs with known organisms suggest that these sequences can be considered as unknown, deep branching clades of Bicosoecida.

Regarding Perkinsea, the reference tree was constructed with 171 sequences from manually selected (Chambouvet *et al*., 2014; Jobard *et al.,* 2020; Itoïz *et al*., unpublished). The model for the phylogenetic reconstruction was GTR + F + R5. In total, 153 ASVs were phylogenetically placed, and the analysis revealed that these sequences were distributed mostly in the NAG01 clade and the environmental clade PERK_ENV02 (Fig. 5C and Fig. S5). Furthermore, one novel ASVs were placed in the Parviluciferaceae group, which is composed mostly of marine dinoflagellates parasites. The great genetic diversity observed represents a potential for novel organisms' isolation and taxonomic clarification of this mostly unknown (and ecologically uncharacterized) group.

We built two phylogenetic trees to classify the novel ASVs related to Mamiellophyceae and Pedinophyceae (Supporting Information Figs. S6 and S7). For Mamiellophyceae, we selected 52 sequences from the bibliography (Lara *et al*., 2017; Marin and Melkonian, 2010). The model used to infer the phylogeny was TN + F + I + G4. The phylogenetic placement of the 53 novel ASVs revealed that many of the sequences, which presented between 93% and 95% of identity with their CCM and CEM (Supporting Information Fig. S6) are sister to Monomastigales (cluster 7). Similar results were observed with the sequences related to Pedinophyceae. For this group, the phylogenetic tree was inferred with 53 sequences selected from the bibliography (Marin, 2012) and with the model TN + F + R4. The clusters 24 and 6, which presented between 92% and 95% of identity with their CIM and CEM, were placed as a sister group related to Pedinomonadales (Supporting Information Fig. S7).

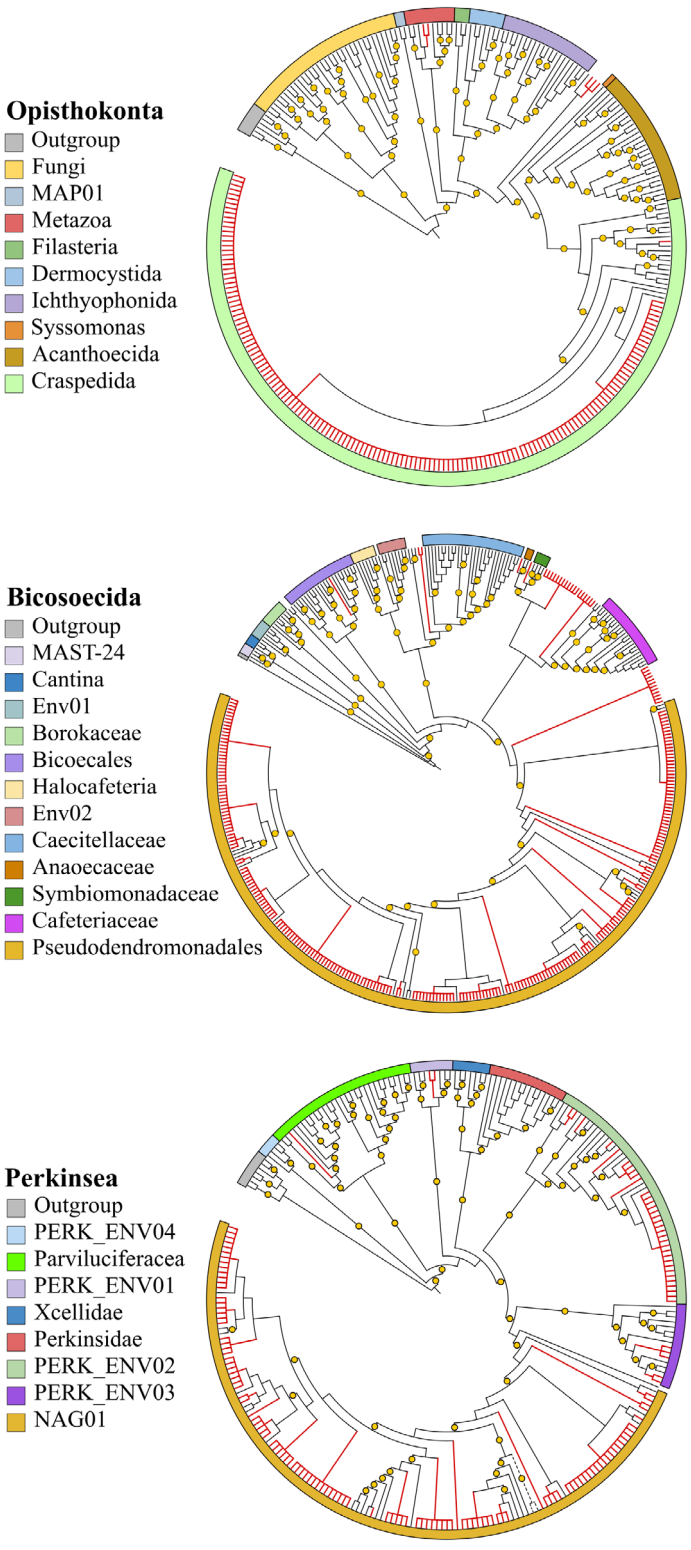

Fig. 5. Phylogeny representation of the protist groups presenting the most novel phylogenetic information in samples from the Paraná fluvial system. The red branches indicate the phylogenetic placement of the ASVs and the yellow dots are bootstrap values >80%. The outside circle is the annotation of the group affiliation. [Color figure can be viewed at wileyonlinelibrary.com]

Discussion

Altogether, we found that ~28% of the total ASVs potentially correspond to novel organisms (i.e. ASVs classified as a novel). Novelty can be characterized both in terms of genetic distances and phylogenetic position of the newly obtained ASVs (del Campo and Massana, 2011;

Lynch *et al*., 2012). Furthermore, as we must differentiate between both novelty relative to identified (barcoded) organisms (*sensu* Pawlowski *et al*., 2012) and in relation to sequences obtained in other environmental surveys, we followed the method proposed by del Campo and Massana (2011). The approach here applied combines all these components of novelty by allowing a rapid screening of each taxonomic group for accurate identification and phylogenetic characterization of the novel groups and new environments by comparing with public and well-documented databases.

Our study reveals a great diversity of novel protists expanding the knowledge from biodiversity in planktonic freshwater communities. So far, few works were focused on the novelty of protist diversity from freshwater. They are related to high mountain lakes (Triadó-Margarit and Casamayor, 2012; Filker *et al*., 2016; Ortiz-Álvarez *et al*., 2018), hypersaline waters (Filker *et al*., 2017) and ancient lakes (Fermani *et al*., 2021), where the novelty can be expected to be subject to a high degree of endemism. Conversely, in the Paraná River, the constant exchange among environments due to the permanent hydrological fluctuations in a spatially heterogeneous floodplain system may enable a high protist diversity. Among freshwater environments, floodplain rivers are considered hotspots of biodiversity (Neiff, 1996; Gopal and Junk, 2000; Amoros and Bornette, 2002). Indeed, the total pool of species found there is derived from the communities inhabiting the variety of habitats arrayed across the main channel and the floodplain (Fournier *et al*., 2020). Three supergroups hosted a major part of taxonomic novelty, Stramenopiles, Alveolata, and Opisthokonta. These supergroups are known to represent an important fraction of the microeukaryotes diversity in freshwater environments (Debroas *et al*., 2017), and our results contribute to broadening that knowledge with the high level of novelty found. In particular, a high diversity of Chrysophyceae, Bacillariophyceae, Dinophyceae, and spirotrichean ciliates are coherent with previous published study in other freshwater systems (Ortiz-Álvarez *et al*., 2018; Singer *et al*., 2021). Remarkably, the Paraná floodplain communities stood out as compared with other freshwater systems due to the high sequence novelty in Bicosoecida (Stramenopiles), Mamiellophyceae and Pedinophyceae (Archaeplastida), Choanoflagellata (Opisthokonta) and Perkinsea (Alveolata), corresponding to deep branching, uncharacterized lineages, highlighting therefore existence of knowledge gaps awaiting to be filled.

The novelty of the ASVs was also explored by their phylogenetic placement showing that many of them are related to environmental clades with long branches, highlighting the presence of novel phylogenetic clades or species. This allows the remarkable finding that recovers the novel ASVs grouped together in the cluster 7 (Supporting Information Fig. S6), related to Mamiellophyceae (Archaeplastida). While the Mamiellales has been mostly studied in marine waters, with many cultures available (Tragin *et al*., 2016), only a few species have been isolated from freshwater (Marin and Melkonian, 2010) although many environmental sequences have been detected (Taib *et al*., 2013; Lara *et al*., 2017). Here, the ASVs from cluster 7 form a deep clade sister to the whole freshwater group Monomastigales; it is therefore likely that these ASVs represent a clade of order rank that has never been detected before. Within green algae, a wealth of novel clusters was also observed within class Pedinophyceae. At the moment only two orders (Marsupiomonadales and Pedinomonadales) and four genera (*Pedinomonas*, *Chlorochytridion*, *Resultor*, and *Marsupiomonas*) have been described (Marin, 2012). Recently, a new phylotype related to *Marsupiomonas* has been retrieved by environmental sequencing (Milyutina *et al*., 2019). Contributing to this finding, the clusters 24 and 6 were related in our study to *Pedinomonas*. These clusters presented a low identity (between 93% and 95%) with publicly available sequences: clusters 24 and 6 formed a deep sister clade to the whole genus. Although the position of clusters 14 (that presented between 90% and 92% of identity with CIM-CEM) within Pedinomonadales is not clear, their genetic similarity with known species suggest that they probably derived from unidentified genera.

Freshwater planktonic systems are thought to be characterized by a higher beta diversity than soil and marine plankton (Singer *et al*., 2021), meaning that they are most dissimilar with each other. For this reason, exploring understudied environments has great chances of revealing uncharted diversity. As was expected, the Paraná fluvial system is, *per se*, highly heterogeneous and contributes in hosting a high microorganism diversity (Devercelli *et al*., 2016; Huber *et al*., 2020). Indeed, our results reveal a great dissimilarity between samples (mean Bray–Curtis distance $>0.78 \pm 0.17$), and a clear separation between samples from the different hydrological periods (Supporting Information Fig. S2). Furthermore, the NMDS revealed the highest similarity in protists community composition during the extreme flood (year 2016) related to the higher connectivity among environments. These samples also present the highest alpha diversity. Similar results were observed in bacterial communities from the same system (Huber *et al*., 2020) and can be explained because the interacting components encompassing this freshwater system (land and water; main channel and floodplain; lentic and lotic water bodies) are mediated by the hydrological dynamic in extreme floods that allows to recruit organisms from

different habitats (Devercelli *et al*., 2014), even from uncharted ones. This complexity brings up a diverse microeukaryote assemblage that may be related to the multiple functional processes that take place in the environment.

Here, we show that non-pigmented organisms represent a large proportion of the diversity of the plankton: indeed, the richness (number of ASVs) of pigmented organisms (~46%) was lower to that of non-pigmented organisms (~51%). The functional analysis of the ASVs showed similar results, where 51% were related to consumers and only 36% were related to phototrophic organisms. The high amount of land-derived organic matter processed in floodplain rivers may contribute to support the outstanding heterotrophic diversity found. Interestingly, the groups that presented the highest novelty were affiliated to non-pigmented organisms, which demonstrates the need for a change of our current vision of freshwater planktonic microbial communities.

In addition, the high diversity of parasitic protists (10% of the total ASVs) may be related to the diversity of micro- and macroorganisms host inhabiting Neotropical environments (Mahé *et al*., 2017). An example is the high diversity of Perkinsea observed, many of them being classified as novel ASVs. This group is exclusively composed so far of parasites infecting a broad range of hosts, from vertebrates, invertebrates and micro-algae (Bråte *et al*., 2010; Mangot *et al*., 2011; Jobard *et al*., 2020). While the ecological importance of the marine counterparts has been well stated (Norén *et al.,* 1999; Burreson *et al.,* 1994; Freeman et al., 2017 ; Itoïz *et al*., unpublished), such as for Perkinsidae, Parviluciferidae and Xcellidae, the role of freshwater Perkinsea remains unknown. Previous studies suggest that they are potential microalgae parasites; however, these results need to be validated with cultures (Mangot *et al*., 2011; Mangot *et al*., 2013; Jobard *et al*., 2020). So far, only one freshwater group, the Severe Perkinsea infection (SPI) agent, was well characterized as a principal parasite responsible for recurrent massive mass mortality of tadpoles and amphibian decline both in temperate (North America and the United Kingdom) and tropical altitudes (Panama, Isidoro-Ayza *et al*., 2017; Smilansky *et al*., 2021). Since these organisms represent a significant number of ASVs in freshwater environments, it is urgent to unveil their diversity and ecological function (Ortiz-Álvarez *et al*., 2018; Salmaso *et al*., 2020; Jobard *et al*., 2020).

Heterotrophic microeukaryotes were represented by the flagellates Bicosoecida and Choanoflagellata with several well-defined new clades (Supporting Information Figs. S3 and S4). In particular, most ASVs affiliated to Bicosoecida were classified within the Pseudodendromonadida, a group formed mainly by freshwater and soil organisms (Cavalier-Smith and Scoble, 2013). Filker *et al*. (2016) also observed a high richness and potential novel diversity in hypersaline waters. This group of small bacterivorous flagellates is common in both marine and freshwater systems (del Campo and Massana, 2011; Massana *et al*., 2021). Although Pseudodendromonadida was considered to be of little ecologically relevance in the past (Massana *et al*., 2007), recent studies provide evidence for a wide distribution of these organisms in marine systems where they grow rapidly in the presence of high bacterial density (Massana *et al*., 2021). In our dataset, this group represented more than 28% of the total ASVs, suggesting that they may play an important role in freshwater ecosystems. Although Choanoflagellata are common in freshwater systems (Singer *et al*., 2021), their immense diversity found was unexpected and especially of organisms related to genus *Codonosiga*. This group has been reported from nearby sites in the Paraná floodplain by Arroyo *et al*. (2018) (as FRESHCH03). Morphological studies in the upper course of the river also found that choanoflagellates and bicosoecids were the most frequent heterotrophic flagellates (Lansac-Toha *et al*., 2016). We suggest that the Paraná fluvial system might be a hotspot for the diversity of these organisms, in line with the high levels of diversity found in Neotropical floodplains; however, further molecular exploration coupled with observational data will be needed to confirm this hypothesis.

Another widely represented heterotrophic group were the ciliates, arguably one of the best documented groups of protists with 27 000–40 000 species described (Lynn, 2008; Gao *et al*., 2016; Weisse, 2017; Clamp and Lynn, 2017; Wang *et al*., 2017) and a well-established and extensive molecular database (Lara and Acosta-Mercado, 2012; Boscaro *et al*., 2018). Still, here we report a large number of ASVs that presented low similarity with any sequences (environmental or cultured/isolated) from the database that could potentially represent an extensive unexplored diversity.

In this work, we applied an approach to characterize the genetic novelty present in an unexplored environment, here the Paraná fluvial system. While different methods have been developed individually, the innovative aspect of this method is the combination of approaches. Indeed, our approach permitted not only characterizing the most novel taxonomic groups but also the existence of clades that are probably still totally unknown. Like in the ocean and in many freshwater systems, the pigmented plankton of the Paraná River has been studied since the 70s (García de Emiliani, 1979, 1981; Devercelli *et al*., 2014). Here, we observed that most of this novel diversity is related to freshwater heterotrophic organisms, encompassing different lifestyles,

from potential heterotrophic predators to parasites and mixotrophs, which are underestimated and poorly surveyed in freshwater research, but which still await to be discovered, isolated, and cultured (if possible).

Experimental procedures

Sampling sites and procedure

In total, 10 samples from the main channel, secondary channels and floodplain lakes were obtained during periods of flood and drought in order to encompass the habitat types and the hydrological variability characteristic of the Paraná fluvial system. The environmental characterization of the sampling sites is presented in the Supporting Information Table S2 and detailed in the study by Mayora *et al*. (2020). Surface water samples for DNA extraction were collected in sterilized bottles and prefiltered through a 50 μm pore mesh to remove invertebrates. The water was filtered through 0.22 μm pore-size filters of mixed cellulose esters (Millipore) until they were clogged with biomass (between 100 and 140 ml) and stored at −80°C until processing.

DNA extraction, PCR and sequencing

Genomic DNA was extracted using a CTAB-based protocol (Zenoff *et al*., 2006). Briefly, filters are incubated at 60°C for 30 min with CTAB lysis buffer, followed by two purification steps with 0.7 ml of chloroform-isoamylic alcohol (24:1) and centrifugation at 14 000 rpm for 10 min. Subsequently, the DNA was precipitated in cold isopropanol and centrifuged again for 30 min. Finally, the pellet was washed with cold ethanol (80%) and the DNA re-suspended into 40 μl of TE buffer.

The V4 region of the 18S rRNA gene was amplified using a nested PCR protocol, with firstly the primers 63f and 1818r (Lepère *et al*., 2011). Followed by a PCR amplification using the TAReuk454FWD1 (Stoeck *et al*., 2010) and V4 18S Next.Rev primers (Piredda *et al*., 2017) for the second round. The PCR conditions and the MiSeq Illumina sequencing were carried out following the protocol previously applied in the study by Metz *et al*. (2019a). Sequences have been deposited at the European Nucleotide Archive (ENA) under the BioProject number (PRJEB24677).

Sequence processing, taxonomic assignation and diversity analysis

For the sequence processing, first we used cutadapt 1.18 (Martin, 2011) to remove the primers. Then, raw sequences were processed with DADA2 V1.10.0 (Callahan *et al*., 2016) with the following parameters for quality filtering: maxEE = (2,2), truncL = (250,220) and truncQ = 2. These parameters were selected according to the quality of the samples. The taxonomic assignment of the ASVs was performed using the Stampa pipeline (Mahé *et al*., 2017), using the PR^2 V4.12 (updated in August 2020) as a reference database (Guillou *et al*., 2013). Those ASVs with less than 10 reads distributed in less than three samples were excluded in order to reduce the possible chimeras and sequencing errors. Furthermore, ASVs assigned to Metazoa and Streptophyta were removed.

The taxonomic composition in terms of number of ASVs (richness) was explored using a treemap plot, performed in R 3.5 (www.r-project.org) and with the package treemapify v 2.5.5 (Wilkins, 2021). The functional annotation of the ASVs was performed following Singer *et al*. (2021). The functional annotation of Dinoflagellates, Cryptophyceae and Chitridyomicotes was manually checked using BLAST (Altschul *et al*., 1997) and assigned according to their similarity to sequences from NCBI database and their phylogenetic position. The alpha diversity (Shannon index) and beta diversity (Bray–Curtis distance) were calculated using the vegan v2.5-7 (Oksanen *et al*., 2008), with the function diversity and *vegandist*, respectively. The NMDS was calculated with the *metaMDS* function from the same R package. For these analyses, the ASV table was rarefacted to the minimum sample size (13 930 reads) with the function rrarefy and for the NMDS the table was then transformed with the Hellinger method using the *deconstand* function.

Analysis of novel diversity

The distribution of the ASVs similarity with PR^2 obtained by Stampa during the taxonomic assignment was analysed to define the supergroups that present a high potential for novel organisms detection using a similar approach proposed by del Campo and Massana (2011) and Lynch *et al*. (2012) described below. The density plot (Fig. 2) was performed using package ggplot2 (Wickham, 2016).

First, all sequences from PR^2 were divided into two reference databases according to their annotation as cultured/isolated or environmental sequences. Using Stampa, we investigate the similarity of each ASV with the CIM and CEM. Then, the similarity among each pair of ASV was calculated by global pairwise alignment as implemented in VSEARCH v2.15.1 (Rognes *et al*., 2016) using the function *--allpairs_global* and min identity of 85% (*--id*), in order to reduce the number of comparisons. ASVs nodes with similarity ≥95% were connected by edges to define clusters of similar ASVs. The scripts for this analysis were programmed with R and are available in GitHub (https://github.com/sebametz/Novelty-SF).

In order to evaluate the degree of novelty, the ASVs were classified according to their similarity with CIM and CEM, which can be interpreted as a degree of knowledge in culture and environmental sequences from public databases, respectively (del Campo and Massana, 2011). We classified the ASVs as 'Novel ASV' if the similarity was smaller than the mean of the supergroup (CIM or CEM for the taxonomic group, quadrant III of the biplot in Fig. 3), 'Environmental ASV' when it presented < mean (CIM) and > mean (CEM) (quadrant IV of the biplot), 'non-novel ASV' (quadrants I and II). Because the mean of the CIM and CEM is associated with each particular supergroup, we defined a global mean (i.e mean of CIM and CEM from the total database) to compare the grade of novelty among supergroups.

Phylogenetic placement of novel ASVs

Phylogenetic analyses were performed for those taxonomic groups where a considerable number of novel ASVs were observed. Briefly, we first performed a reference tree with sequences retrieved from NCBI. The sequences were aligned using MAFFT version 7.453 (Katoh and Standley, 2013) with the E-INS-i algorithm and trimmed with TrimAl (Capella-Gutiérrez *et al*., 2009) with parameter -gt 0.6; and the phylogeny was reconstructed using IQ-Tree2 (Nguyen *et al.,* 2015; Minh *et al*., 2020) with parameters --altr 1000, for SH-aLRT test (Guindon *et al*., 2010), and --m MF, for automatic model detection with ModelFinder (Kalyaanamoorthy *et al*., 2017). The tree was annotated according to the bibliography and the bootstrap values. Following the references tree obtaining we classified the ASVs using a placement pipeline, where the sequences were aligned to the reference alignment using MAFFT with --addfragment and --keeplength parameters and placed in the phylogenetic tree using EPA-ng v0.3.8 (Barbera *et al*., 2019). The results were converted from a .jplace file to a .nexwik file using GAPPA (Czech *et al*., 2020) with the *examine graft* function. The final phylogenetic tree was visualized and annotated using Itol (https://itol.embl.de/, Letunic and Bork, 2019) and edited with Inkscape (ww.w.inkscape.org). The alignments and the phylogenetic trees are available in GitHub (https://github.com/sebametz/Novelty-SF).

Acknowledgements

This study was supported by Agencia Nacional de Promoción Científica y Tecnológica (PICT 2016-0465, PI: MD), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). S.M. is supported by the Région Bretagne fellowship (SAD 2019 'PlastPerk' N°1537). The authors thank C. Debonis, E. Creus, and M. Piacenza for their field assistance, E. Rueda for helping in DNA extractions, and A. Dabin for bioinformatic support. Bioinformatics analyses were performed at the PIRAYU cluster (https://cimec.org.ar/c3/pirayu/index.php). E.L. was funded by a project 'Atracción de talento investigador' by the Consejería de Educación, Juventud y Deporte, Comunidad de Madrid (Spain) 2017-T1/AMB-5210 and by and the project MYXOTROPIC VI (PGC2018-094660-B-I00) from the Spanish Government.

References

Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., *et al*. (2019) Revisions to the classification, nomenclature, and diversity of eukaryotes. *J Euk Microb* **66**: 4–119.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Amoros, C., and Bornette, G. (2002) Connectivity and biocomplexity in waterbodies of riverine floodplains. *Fresh Biol* **47**: 761–776.

Annenkova, N.V., Giner, C.R., and Logares, R. (2020) Tracing the origin of planktonic protists in an ancient lake. *Microorganisms* **8**: 543.

Arroyo, A.S., Iannes, R., Bapteste, E., and Ruiz-Trillo, I. (2020) Gene similarity networks unveil a potential novel unicellular group closely related to animals from the Tara oceans expedition. *Genome Biol Evol* **12**: 1664–1678.

Arroyo, A.S., López-Escardó, D., Kim, E., Ruiz-Trillo, I., and Najle, S.R. (2018) Novel diversity of deeply branching Holomycota and unicellular holozoans revealed by metabarcoding in middle Paraná River, Argentina. *Front Ecol Evol* **6**: 99.

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.

Bjorbækmo, M.F.M., Evenstad, A., Røsæg, L.L., Krabberød, A.K., and Logares, R. (2020) The planktonic protist interactome: where do we stand after a century of research? *ISME J* **14**: 544–559.

Bock, C., Jensen, M., Forster, D., Marks, S., Nuy, J., Psenner, R., *et al*. (2020) Factors shaping community patterns of protists and bacteria on a European scale. *Environ Microb* **22**: 2243–2260.

Boscaro, V., Santoferrara, L., Zhang, Q., Gentekaki, E., Syberg-Olsen, M., del Campo, J., and Keeling, P. (2018) EukRef–Ciliophora: a manually curated, phylogeny-based database of small subunit rRNA gene sequences of ciliates. *Environ Microbiol* **20**: 2218–2230.

Bråte, 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. (2020) The new tree of eukaryotes. *Trends Ecol Evol* **35**: 43–55.

Burreson, E.M., Alvarez, R.S., Martinez, V.V., and Macedo, L.A. (1994). Perkinsus marinus (Apicomplexa) as a potential source of oyster Crassostrea virginica mortality in coastal lagoons of Tabasco, Mexico. *Diseases of Aquatic Organisms* **20**: 77–82. https://doi.org/10.3354/dao020077

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. *Nat Methods* **13**: 581–583.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, 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.

Cavalier-Smith, T., and Scoble, J.M. (2013) Phylogeny of Heterokonta: *Incisomonas marina*, a uniciliate gliding opalozoan related to *Solenicola* (Nanomonadea), and evidence that Actinophryida evolved from Raphidophytes. *Eur J Prot* **49**: 328–353.

Chambouvet, A., Berney, C., Romac, S., Audic, S., Maguire, F., De Vargas, C., and Richards, T.A. (2014) Diverse molecular signatures for ribosomally 'active' Perkinsea in marine sediments. *BMC Microbiol* **14**: 1–10.

Clamp, J., and Lynn, D. (2017) Investigating the biodiversity of ciliates in the 'Age of Integration'. *Eur J Protistol* **61**: 314–322.

Corlett, R.T. (2017) A bigger toolbox: biotechnology in biodiversity conservation. *Trends Biotechnol* **35**: 55–65.

Czech, L., Barbera, P., and Stamatakis, A. (2020) Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics* **36**: 3263–3265.

Debroas, D., Domaizon, I., Humbert, J.F., Jardillier, L., Lepère, C., Oudart, A., and Taïb, N. (2017) Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available metabarcoding data. *FEMS Microb Ecol* **93**: fix023.

del Campo, J., Kolisko, M., Boscaro, V., Santoferrara, L.F., Nenarokov, S., Massana, R., *et al*. (2018) EukRef: phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. *PLoS Biol* **16**: e2005849.

del Campo, J., and Massana, R. (2011) Emerging diversity within Chrysophytes, Choanoflagellates and Bicosoecids based on molecular surveys. *Protist* **162**: 435–448.

Devercelli, M., de Domitrovic, Y.Z., Forastier, M.E., and de Zaburlín, N.M. (2014) Phytoplankton of the Paraná River basin. *Adv Limnol* **65**: 39–65.

Devercelli, M., Scarabotti, P., Mayora, G., Schneider, B., and Giri, F. (2016) Unravelling the role of determinism and stochasticity in structuring the phytoplanktonic metacommunity of the Paraná River floodplain. *Hydrobiologia* **764**: 139–156.

Dunthorn, M., Otto, J., Berger, S.A., Stamatakis, A., Mahé, F., Romac, S., *et al*. (2014) Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. *Mol Biol Evol* **31**: 993–1009.

Falkowski, P., Laws, E., Barber, R., and Murray, J. (2003) Phytoplankton and their role in primary, new, and export production. In *Ocean Biogeochemistry. Global Change — The IGBP Series*, Fasham, M.J.R. (ed). Berlin: Springer, pp. 99–121.

Fermani, P., Metz, S., Balagué, V., Descy, J.P., Morana, C., Logares, R., *et al*. (2021) Microbial eukaryote assemblages and potential novel diversity in four tropical east African Great Lakes. *FEMS Microbiol Ecol* **97**: fiab114.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**: 237–240. https://doi.org/10.1126/science.281.5374.237

Filker, S., Forster, D., Weinisch, L., Mora-Ruiz, M., González, B., Farías, M.E., *et al*. (2017) Transition boundaries for protistan species turnover in hypersaline waters of different biogeographic regions. *Environ Microbiol* **19**: 3186–3200.

Filker, S., Gimmler, A., Dunthorn, M., Mahé, F., and Stoeck, T. (2015) Deep sequencing uncovers protistan plankton diversity in the Portuguese ria Formosa solar saltern ponds. *Extremophiles* **19**: 283–295.

Filker, S., Sommaruga, R., Vila, I., and Stoeck, T. (2016) Microbial eukaryote plankton communities of high-mountain lakes from three continents exhibit strong biogeographic patterns. *Mol Ecol* **25**: 2286–2301.

Fournier, B., Samaritani, E., Frey, B., Seppey, C.V., Lara, E., Heger, T.J., and Mitchell, E.A. (2020) Higher spatial than seasonal variation in floodplain soil eukaryotic microbial communities. *Soil Biol Biochem* **147**: 107842.

Freeman, M.A., Fuss, J., Kristmundsson, Á., Bjorbækmo, Marit F.M., Mangot, J.-F., del Campo, J., Keeling, P.J., Shalchian-Tabrizi, K., Bass, D. (2017). X-Cells Are Globally Distributed, Genetically Divergent Fish Parasites Related to Perkinsids and Dinoflagellates. *Current Biology* **27**: 1645–1651.e3. https://doi.org/10.1016/j.cub.2017.04.045

Gao, F., Warren, A., Zhang, Q., Gong, J., Miao, M., Sun, P., *et al*. (2016) The all-data-based evolutionary hypothesis of ciliated protists with a revised classification of the phylum Ciliophora (Eukaryota, Alveolata). *SC Rep* **6**: 24874.

Gao, Z., Karlsson, I., Geisen, S., Kowalchuk, G., and Jousset, A. (2019) Protists: puppet masters of the rhizosphere microbiome. *Trends Plant Sci* **24**: 165–176.

García de Emiliani, M.O. (1979) Campaña "Keratella I" a lo Largo del Río Paraná Medio. III: Fitoplancton de Ambientes Lenticos. *Natura Neotropicalis* **1**: 73–84.

García de Emiliani, M.O. (1981) Fitoplancton de los Principales Cauces y Tributarios del Valle Aluvial del Río Paraná: Tramo Goya-Diamante. *Natura Neotropicalis* **1**: 112–125.

Geisen, S., Mitchell, E.A., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., *et al*. (2018) Soil protists: a fertile frontier in soil biology research. *FEMS Microb Rev* **42**: 293–323.

Gopal, B., and Junk, W.J. (2000) Biodiversity in wetlands: an introduction. In *Biodiversity in Wetlands: Assessment, Function and Conservation*, Vol. **1**. Leiden: Backhuys Publishers, pp. 1–10.

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., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307–321.

Huber, P., Metz, S., Unrein, F., Mayora, G., Sarmento, H., and Devercelli, M. (2020) Environmental heterogeneity determines the ecological processes that govern bacterial metacommunity assembly in a floodplain river system. *ISME J* **14**: 2951–2966.

Isidoro-Ayza, M., Lorch, J.M., Grear, D.A., Winzeler, M., Calhoun, D.L., and Barichivich, W.J. (2017) Pathogenic lineage of Perkinsea associated with mass mortality of frogs across the United States. *Sci Rep* **7**: 1–10.

Jobard, M., Wawrzyniak, I., Bronner, G., Marie, D., Vellet, A., Sime-Ngando, T., *et al*. (2020) Freshwater Perkinsea: diversity, ecology and genomic information. *J Plankton Res* **42**: 3–17.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., Von Haeseler, A., and Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**: 587–589.

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.

Keeling, P.J., and del Campo, J. (2017) Marine protists are not just big bacteria. *Curr Biol* **27**: R541–R549.

Lansac-Tôha, F.M., Meira, B.R., Segovia, B.T., Lansac-Tôha, F.A., and Velho, L.F.M. (2016) Hydrological connectivity determining metacommunity structure of planktonic heterotrophic flagellates. *Hydrobiologia* **781**: 81–94.

Lara, E., and Acosta-Mercado, D. (2012) A molecular perspective on ciliates as soil bioindicators. *Eur J Soil Biol* **49**: 107–111.

Lara, E., Fernández, L.D., Schiaffino, M.R., and Izaguirre, I. (2017) First freshwater member ever reported for the family Bathycoccaceae (Chlorophyta; Archaeplastida) from Argentinean Patagonia revealed by environmental DNA survey. *Europ J Protist* **60**: 45–49.

Lara, E., Moreira, D., and López-García, P. (2010) The environmental clade LKM11 and *Rozella* form the deepest branching clade of fungi. *Protist* **161**: 116–121.

Lentendu, G., Buosi, P.R.B., Cabral, A.F., Trevizan Segóvia, B., Ramos Meira, B., Lansac-Tôha, F.M., *et al*. (2018) Protist biodiversity and biogeography in lakes from four Brazilian river–floodplain systems. *J Euk Microb* **66**: 592–599.

Lepère, C., Demura, M., Kawachi, M., Romac, S., Probert, I., and Vaulot, D. (2011) Whole-genome amplification (WGA) of marine photosynthetic eukaryote populations. *FEMS Microb Ecol* **76**: 513–523.

Letunic, I., and Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**: 256–259.

Lynch, M.D., Bartram, A.K., and Neufeld, J.D. (2012) Targeted recovery of novel phylogenetic diversity from next-generation sequence data. *ISME J* **6**: 2067–2077.

Lynn, D. (2008) *The Ciliated Protozoa: Characterization, Classification, and Guide to the Literature*. Dordrecht: Springer.

Machado, K.B., Targueta, C.P., Antunes, A.M., Soares, T.N., Telles, M.P.D.C., Logares, R., *et al*. (2019) Diversity patterns of planktonic microeukaryote communities in tropical floodplain lakes based on 18S rDNA gene sequences. *J Plankton Res* **41**: 241–256.

Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., *et al*. (2017) Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat Ecol Evol* **1**: 0091.

Mangot, J.F., Debroas, D., and Domaizon, I. (2011) Perkinsozoa, a well-known marine protozoan flagellate parasite group, newly identified in lacustrine systems: a review. *Hydrobiologia* **659**: 37–48.

Mangot, J.F., Domaizon, I., Taib, N., Marouni, N., Duffaud, E., Bronner, G., and Debroas, D. (2013) Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. *Environ Microb* **15**: 1745–1758.

Marin, B. (2012) Nested in the Chlorellales or independent class? Phylogeny and classification of the Pedinophyceae (Viridiplantae) revealed by molecular phylogenetic analyses of complete nuclear and plastid-encoded rRNA operons. *Protist* **163**: 778–805.

Marin, B., and Melkonian, M. (2010) Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear-and plastid-encoded rRNA operons. *Protist* **161**: 304–336.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**: 10–12.

Massana, R., Castresana, J., Balagué, V., Guillou, L., Romari, K., Groisillier, A., *et al*. (2004) Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl Environ Microb* **70**: 3528–3534.

Massana, R., del Campo, J., Dinter, C., and Sommaruga, R. (2007) Crash of a population of the marine heterotrophic flagellate *cafeteria roenbergensis* by viral infection. *Environ Microb* **9**: 2660–2669.

Massana, R., Labarre, A., López-Escardó, D., Obiol, A., Bucchini, F., Hackl, T., *et al*. (2021) Gene expression during bacterivorous growth of a widespread marine heterotrophic flagellate. *ISME J* **15**: 154–167.

Massana, R., Pernice, M., Bunge, J.A., and del Campo, J. (2011) Sequence diversity and novelty of natural assemblages of picoeukaryotes from the Indian Ocean. *ISME J* **5**: 184–195.

Matsen, F.A., Kodner, R.B., and Armbrust, E.V. (2010) pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**: 538.

Mayora, G., Scarabotti, P., Schneider, B., Alvarenga, P., and Marchese, M. (2020) Multiscale environmental heterogeneity in a large river floodplain system. *J South Am Earth Sci* **100**: 102546.

Metz, S., Lopes dos Santos, A., Berman, M.C., Bigeard, E., Licursi, M., Not, F., *et al*. (2019a) Diversity of photosynthetic picoeukaryotes in eutrophic shallow lakes as assessed by combining flow cytometry cell-sorting and high throughput sequencing. *FEMS Microb Ecol* **95**: fiz038.

Metz, S., Singer, D., Domaizon, I., Unrein, F., and Lara, E. (2019b) Global distribution of Trebouxiophyceae diversity explored by high-throughput sequencing and phylogenetic approaches. *Environ Microb* **21**: 3885–3895.

Milyutina, I.A., Belevich, T.A., Ilyash, L.V., and Troitsky, A.V. (2019) Insight into picophytoplankton diversity of the subarctic White Sea—the first recording of Pedinophyceae in environmental DNA. *Microbiol Open* **8**: e892.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* **37**: 1530–1534.

Moon-van der Staay, S.Y., De Wachter, R., and Vaulot, D. (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409**: 607–610.

Moreira, D., and López-García, P. (2014) The rise and fall of Picobiliphytes: how assumed autotrophs turned out to be heterotrophs. *Bioessays* **36**: 468–474.

Neiff, J. (1996) Large rivers of South America: toward the new approach. *Verh Int Verein Limnol* **26**: 167–180.

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.

Norén, F., Moestrup, Ø., and Rehnstam-Holm, A.S. (1999) *Parvilucifera infectans* Norén et Moestrup gen. et sp. nov. (Perkinsozoa *phylum* nov.): a parasitic flagellate capable of killing toxic microalgae. *Eur J Protist* **35**: 233–254.

Not, F., Gausling, R., Azam, F., Heidelberg, J.F., and Worden, A.Z. (2007) Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environ Microb* **9**: 1233–1252.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G.L., Stevens, M.H.H., and Wagner, H. (2008) The vegan package: community ecology package, version 1.13-1 URL http://vegan.r-forge.r-project.Org

Ortiz-Álvarez, R., Triadó-Margarit, X., Camarero, L., Casamayor, E.O., and Catalan, J. (2018) High planktonic diversity in mountain lakes contains similar contributions of autotrophic, heterotrophic and parasitic eukaryotic life forms. *Sci Rep* **8**: 1–11.

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., *et al*. (2012) CBOL Protist Working Group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol* **10**: e1001419.

Pedrós-Alió, C., Acinas, S.G., Logares, R., and Massana, R. (2018) Marine microbial diversity as seen by high throughput sequencing. In *Microbial Ecology of the Oceans*, Gasol, J.M., and Kirchman, D.L. (eds). New York, NY: John Wiley and Sons, pp. 47–98.

Piredda, R., Tomasino, M.P., D'erchia, A.M., Manzari, C., Pesole, G., Montresor, M., *et al*. (2017) Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean long term ecological research site. *FEMS Microb Ecol* **93**: fiw200.

Roberto, M.C., Santana, N.F., and Thomaz, S.M. (2009) Limnology in the upper Paraná River floodplain: large-scale spatial and temporal patterns, and the influence of reservoirs. *Brazil J Biol* **69**: 717–725.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**: e2584.

Salmaso, N., Boscaini, A., and Pindo, M. (2020) Unraveling the diversity of eukaryotic microplankton in a large and deep perialpine lake using a high throughput sequencing approach. *Front Microb* **11**: 789.

Schoenle, A., Hohlfeld, M., Rosse, M., Filz, P., Wylezich, C., Nitsche, F., and Arndt, H. (2020) Global comparison of bicosoecid *Cafeteria*-like flagellates from the deep ocean and surface waters, with reorganization of the family Cafeteriaceae. *Eur J Protistol* **73**: 125665.

Seeleuthner, Y., Mondy, S., Lombard, V., Carradec, Q., Pelletier, E., Wessner, M., *et al*. (2018) Single-cell genomics of multiple uncultured stramenopiles reveals underestimated functional diversity across oceans. *Nat Commun* **9**: 1–10.

Seenivasan, R., Sausen, N., Medlin, L.K., and Melkonian, M. (2013) Picomonas judraskeda gen. et sp. nov.: the first identified member of the Picozoa phylum nov., a widespread group of picoeukaryotes, formerly known as 'picobiliphytes'. *PLoS one* **8**: e59565.

Singer, D., Seppey, C.V.W., Lentendu, L., Micah, D., Bass, D., Belbahri, L., *et al*. (2021) Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Environ Int* **146**: 106262.

Smilansky, V., Jirků, M., Milner, D.S., Ibáñez, R., Gratwicke, B., Nicholls, A., *et al*. (2021) Expanded host and geographic range of tadpole associations with the severe Perkinsea infection group. *Biol Lett* **17**: 20210166.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D., Breiner, H.W., and Richards, T.A. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* **19**: 21–31.

Strom, S.L. (2008) Microbial ecology of ocean biogeochemistry: a community perspective. *Science* **320**: 1043–1045.

Taib, N., Mangot, J.F., Domaizon, I., Bronner, G., and Debroas, D. (2013) Phylogenetic affiliation of SSU rRNA genes generated by massively parallel sequencing: new insights into the freshwater protist diversity. *PLoS One* **8**: e58950.

Torruella, G., Grau-Bové, X., Moreira, D., Karpov, S.A., Burns, J.A., Sebé-Pedrós, A., *et al*. (2018) Global transcriptome analysis of the aphelid *Paraphelidium tribonemae* supports the phagotrophic origin of fungi. *Commun Biol* **1**: 1–11.

Tragin, M., Lopes dos Santos, A., Christen, R., and Vaulot, D. (2016) Diversity and ecology of green microalgae in marine systems: an overview based on 18S rRNA gene sequences. *Perspect Phycol* **3**: 141–154.

Triadó-Margarit, X., and Casamayor, E.O. (2012) Genetic diversity of planktonic eukaryotes in high mountain lakes (Central Pyrenees, Spain). *Environ Microbiol* **14**: 2445–2456.

Wang, P., Wang, Y., Wang, C., Zhang, T., Alfarraj, S., and Gao, F. (2017) Further consideration on the phylogeny of the Ciliophora: analyses using both mitochondrial and nuclear data with focus on the extremely confused class Phyllopharyngea. *Mol Phyl Evol* **112**: 96–106.

Weisse, T. (2017) Functional diversity of aquatic ciliates. *Eur J Protistol* **61**: 331–358.
Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Berlin: Springer.
Wilkins, D. (2021). treemapify: draw Treemaps in ''ggplot2.'' R package version 2.5.5.
Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E., and Keeling, P.J. (2015) Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347**: 1257594.
Zenoff, V.F., Siñeriz, F., and Farias, M.E. (2006) Diverse responses to UV-B radiation and repair mechanisms of bacteria isolated from high-altitude aquatic environments. *Appl Environ Microb* **72**: 7857–7863.

Supporting Information

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

Fig. S1. a) Location of the Paraná fluvial system in South America showing the studied area, and b) the sampled environments.

Fig. S2. Ordination plot (non-metric multidimensional scaling) of ASVs protists communities from the different years, corresponding to different hydrological periods. In yellow samples from 2014, related to a drought period. In green samples from 2015, related to flood period and in blue samples from 2016, related to extreme drought period.

Fig. S3. Opisthokonta phylogenetic tree and novel ASVs related to Choanoflagellata placement. Bootstrap values >80% are shown by yellow dots. Red branches correspond to the placement of the ASVs from the sequence similarity analysis (Fig. 4). The NCBI accession number of the reference sequences are annotated in the tree leaf. The squares correspond to the annotation of the group affiliation.

Fig. S4. Bicosoecida phylogenetic tree and novel ASVs classification. Bootstrap values >80% are shown by yellow dots. Red branches correspond to the placement of the ASVs from the sequence similarity analysis (Fig. 4). The NCBI accession number of the reference sequences are annotated in the tree leaf. The squares correspond to the annotation of the group affiliation.

Fig. S5. Perkinsea phylogenetic tree and novel ASVs classification. Bootstrap values >80% are shown by yellow dots. Red branches correspond to the placement of the ASVs from the sequence similarity analysis (Fig. 4). The NCBI accession number of the reference sequences are annotated in the tree leaf. The squares correspond to the annotation of the group affiliation.

Fig. S6. Mamiellophyceae phylogenetic tree and novel ASVs classification. Bootstrap values >80% are shown by yellow dots. Red branches correspond to the placement of the ASVs from the sequence similarity analysis (Fig. 4). The NCBI accession number of the reference sequences are annotated in the tree leaf. The squares correspond to the annotation of the group affiliation.

Fig. S7. 'Core' Archaeplastida phylogenetic tree for Pedinophyceae novel ASVs classification. Bootstrap values >80% are shown by yellow dots. Red branches correspond to the placement of the ASVs from the sequence similarity analysis (Fig. 4). The NCBI accession number of the reference sequences are annotated in the tree leaf. The squares correspond to the annotation of the group affiliation.

Table S1 Contribution of each taxonomic group to total freshwater protists richness studied with 18S rDNA sequencing in the Paraná fluvial system. The classes correspond to those defined by the PR^2 database. The * correspond to ASVs with unknown class classification.

Table S2. Freshwater environments studied within the Paraná fluvial system and environmental characteristics. The river's hydrological condition (drought, flood, and extreme flood) and habitat type of each environment are indicated; lakes and ponds are considered isolated or connected if they were or not hydrologically connected to a lotic water body. Cond: conductivity, DO: dissolved oxygen, TN: total nitrogen, TP: total phosphorus, Chl-a: chlorophyll-a.

Table S3. ASVs classification according to if the related classification corresponds to a pigmented or non-pigmented protists at the Paraná fluvial system.

Table S4. Protist diversity at the Paraná fluvial system: Alpha diversity of each sample computed with Shannon index, and Beta diversity among samples calculated with Bray-Curtis distance. In red are highlighted distances between samples lower than 0.5.