<http://mc.manuscriptcentral.com/fems>

Microbial eukaryotes assemblages and potential novel diversity in four tropical East-African Great Lakes

Journal:	<i>FEMS Microbiology Ecology</i>
Manuscript ID	FEMSEC-21-04-0167.R1
Manuscript Type:	Research article
Date Submitted by the Author:	n/a
Complete List of Authors:	Fermani, Paulina; Instituto Tecnológico de Chascomús, INTECH Metz, Sebastian; Instituto de Investigaciones Biotecnológicas-Instituto Tecnológico de Chascomús, IIB-INTECH Balague, Vanessa; ICM Barcelona, CSIC Descy, Jean-Pierre; , Université de Namur, Laboratory of Freshwater Ecology Morana, Cédric; Université de Liège, Chemical Oceanography Unit Logares, Ramiro; Institut de Ciències del Mar, Department of Marine Biology Massana, Ramon; Institut de Ciències del Mar, CSIC, Biologia Marina i Oceanografia Sarmiento, Hugo; Universidade Federal de Sao Carlos, Hidrobiologia
Keywords:	Protists, metabarcoding, diversity, novelty, great lakes, Africa

SCHOLARONE™
Manuscripts

Microbial Eukaryote Diversity in African Lakes

Manuscript submitted to FEMS Microbiology Ecology

Microbial eukaryotes assemblages and potential novel diversity in four tropical East-African Great Lakes

Paulina Fermani^{1*}, Sebastián Metz², Vanessa Balagué³, Jean Pierre Descy⁴, Cédric Morana⁴, Ramiro Logares³, Ramon Massana³, Hugo Sarmento⁵

¹Laboratorio de Ecología Acuática. Instituto Tecnológico de Chascomús (UNSAM-CONICET) Chascomús, Buenos Aires, Argentina.

²University of Brest, CNRS, IRD, Ifremer, LEMAR, F-29280, Plouzané, France

³Institut de Ciències del Mar (CSIC), Passeig Marítim de la Barceloneta, 37-49, ES-08003, Barcelona, Catalonia, Spain.

⁴Unit of Chemical Oceanography, University of Liège, Liège, Belgium.

⁵Departamento de Hidrobiologia (DHB), Universidade Federal de São Carlos, São Carlos - SP, Brazil.

***Corresponding author:**

Paulina Fermani. Instituto Tecnológico de Chascomús (INTECH-UNSAM-CONICET), Av. Intendente Marino Km 8.2, Chascomús (7130), Buenos Aires, Argentina.

e-mail: pfermani@intech.gov.ar

Phone (+54) 2241-430323, Ext110

Keywords: Protists; metabarcoding; diversity; novelty; great lakes; Africa

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract

East-African Great Lakes are old and unique natural resources heavily utilized by their bordering countries. In those lakes, ecosystem functioning is dominated by pelagic processes, where microorganisms are key components, however protistan diversity is barely known. We investigated the community composition of small eukaryotes (< 10 µm) in surface waters of four African Lakes (Kivu, Edward, Albert and Victoria) by sequencing the 18S rRNA gene. Moreover, in the meromictic Lake Kivu, two stations were vertically studied. We found high protistan diversity distributed in 779 operational taxonomic units (OTUs), spanning in eleven high-rank lineages, being Alveolata (31%), Opisthokonta (20%) and Stramenopiles (17%) the most represented supergroups. Surface protistan assemblage were associated to conductivity and productivity gradients; whereas depth, had a strong effect on protistan community in Kivu, with higher contribution of heterotrophic organisms. Approximately 40% of OTUs had low similarity (< 90%) with reported sequences in public databases, these were mostly coming from deep anoxic waters of Kivu, suggesting a high extent of novel diversity. We also detected several taxa so far considered exclusive of marine ecosystems. Our results unveiled a complex and largely undescribed protistan community, in which several lineages have adapted to different niches after crossing the salinity boundary.

Introduction

The East African Rift valley harbors some of the world largest lakes, such as Lake Tanganyika, Malawi, Edward, Albert, Kivu, and Victoria, the second largest freshwater in the world (Ogutu-Ohwayo *et al.* 1997). These African Great Lakes provide essential ecosystem services to their bordering populations, like water supply, fisheries, recreation and tourism (Cohen *et al.* 1996; Odada *et al.* 2003). Some of them are among the oldest freshwater systems in the world, displaying an array of limnological and biological features, ranging from productive waters to extremely oligotrophic systems (Bootsma and Hecky 1993; Loiselle *et al.* 2014; İnceoğlu *et al.* 2015). They have attracted biologists from the beginning of the 19th century for their very large vertebrate and invertebrate biodiversity, with for instance hundreds of endemic fish species (e.g. Brooks 1950; Fryer and Iles 1972; Seehausen *et al.* 1997). Large tropical lakes differ from the temperate ones mainly by the so-called 'endless summer' (Kilham and Kilham 1990), with high temperatures throughout the year. Due to their large size, ecosystem function is dominated by pelagic processes, where microorganisms are key components (Descy and Sarmiento 2008). However, the high human population density inhabiting near the coast generates high resource exploitation, pollution, species introduction, and eutrophication, which together with climate change impacts, alter lake biological diversity and function (Hecky and Bugenyi 1992; Bootsma and Hecky 1993; Otu *et al.* 2011).

Although they share common characteristics, the African Great lakes display a variety of limnological conditions, which have to be taken into account for understanding their functioning and predicting their future, including their response to climate and other anthropogenic influences (Bootsma and Hecky 1993; Odada *et al.* 2003). In general, environmental heterogeneity produces community differences in waters bodies, important structuring abiotic factors on the local scale being temperature, light, depth, pH, conductivity, nutrients, among others (Wu *et al.* 2009; Weisse *et al.* 2016). In particular, the deepest lakes (Tanganyika, Malawi and Kivu) are meromictic. As opposed to the shallower lakes (Victoria, Edward and Albert), which are holomictic, they never experience complete mixing of the water column, due to their great depth and the existence of a salinity gradient, separating the oxic surface waters (the mixolimnion) from the permanently anoxic deep waters (the monimolimnion). Because they have a chemocline with a redox gradient at the transition between the

1
2
3 86 surface and deep waters, meromictic lakes have many ecological niches for microbial communities
4
5 87 (Oikonomou *et al.* 2014; Lepère *et al.* 2016). The anoxic and saline waters below the surface layers
6
7 88 create another extreme environment (Charvet *et al.* 2012).
8

9 89 Lake Kivu is of particular interest because the vertical mixing and transport processes are
10
11 90 different from most other large lakes in the world (Schmid and Wüest 2012). The oxic mixolimnion
12
13 91 responds to the same atmospheric forcing as the other Rift lakes, with a relatively weak thermal
14
15 92 stratification. On the other hand, the deep waters display a unique structure because they receive heat,
16
17 93 salts, and CO₂ from deep geothermal inflows, which increase the salinity, nutrients and methane
18
19 94 concentrations (Schmid *et al.* 2002; Borges *et al.* 2011). The lake also shows significant spatial
20
21 95 heterogeneity, as differences in the vertical structure of the water column between the Northern and
22
23 96 Southern basins have been observed (Morana *et al.* 2015; Roland *et al.* 2017).
24
25

26 97 Single-celled eukaryotes (protists) are key for ecosystem functioning (Sherr and Sherr 1998).
27
28 98 Autotrophs are the main carbon fixers in aquatic environments, whereas heterotrophs promote nutrient
29
30 99 cycling by consuming bacteria and picophytoplankton, being also crucial in determining the transport
31
32 100 of organic matter to higher trophic levels (Pomeroy 1974; Boenigk and Arndt 2002; Stenuite *et al.*
33
34 101 2009). Moreover, parasitic protists influence community dynamics of larger eukaryotic hosts (Worden
35
36 102 *et al.* 2015). Several authors have observed that carbon transfer through this microbial food web to
37
38 103 higher trophic levels is significantly more important in oligotrophic than eutrophic systems (Porter *et*
39
40 104 *al.* 1988; Weisse *et al.* 1990; Domingues *et al.* 2017), as suspected by Hecky *et al.* (1981) and
41
42 105 demonstrated by Pirlot *et al.* (2005) and Tarbe *et al.* (2011a) in Lake Tanganyika and Kivu (Sarmiento
43
44 106 *et al.* 2008). Nevertheless, due to the large number of interactions between planktonic components in
45
46 107 the most productive systems, this trend is still under debate (Fermani *et al.* 2015; Xiong *et al.* 2021).
47
48

49 108 Over the last decades, studies of eukaryotic microbial communities in freshwater systems
50
51 109 through molecular analysis have increased (Simon *et al.* 2014; Schloss *et al.* 2016; Debroas *et al.*
52
53 110 2017; Annenkova *et al.* 2020). Recently, two analyses across diverse habitats including freshwater,
54
55 111 marine and soils, revealed clear differences in the taxonomic composition of the major protistan
56
57 112 lineages as well as a higher β -diversity in freshwater than in the other systems (Singer *et al.* 2021;
58
59 113 Xiong *et al.* 2021). Despite their crucial ecological roles, protists assemblages is barely known in
60

African lakes (e.g. Tarbe *et al.* 2011b; Llíros *et al.* 2012), although they harbor an important part of the world biodiversity heritage (see e.g. Vadeboncoeur *et al.* 2011), and habitats with oxygen gradients has received very little attention (e.g Tarbe *et al.* 2011; Oikonomou *et al.* 2014; Lèpere *et al.* 2016).

Taking account their importance in food webs and in order to fill these gaps, our first aim was to describe the small eukaryote (< 10 µm) assemblage in four East African Great Lakes (Edward, Albert, Victoria and Kivu) using high-throughput amplicon tag pyrosequencing of the V4 region of the 18S rRNA gene. Because of their ancestry, these lakes might be able to support a high rate of endemism (e.g. Cichlids), therefore we expect to find a high degree of potential novel diversity. Further, our second goal was to compare the surface layers of Lakes Edward, Albert, Victoria and two basins of Kivu (the main basin, which is the most exposed to the winds, and the southern basin, which is more wind-protected) and to determine the main drivers shaping the structure of communities. Although these freshwater systems share common characteristics (same geological formation, tropical, large and old lakes), we hypothesize that their protist community may differ, due to morphological, physical and chemical differences among lakes. Moreover, considering the unique characteristics of Lake Kivu, our third objective was to compare the eukaryote microbial composition over a vertical profile in the redox gradient at two sites: Gisenyi (northern basin) and Ishungu (southern basin). We hypothesize that both sites exhibit a dissimilar protistan assemblage, due to the physical-chemical differences of the redox gradient. In turn, we expect to find a high and barely known diversity, especially in the deepest areas where the conditions are extreme.

Material and Methods

Study sites

East African Great Lakes are located between 4° 35' N and 14° 30' S of latitude, covering 2,100 km north-south distance (Spigel and Coulter 1996). The region has a tropical climate with a bimodal distribution in rainfall, which defines two seasons: a dry season with south-eastern dominant winds (from June to September) and a rainy season (from October to May), calmer and warmer (Descy and Sarmiento 2008). Some of the main lakes in the region are Kivu, Edward, Albert and

Victoria (Fig. 1A). Each one is situated in a different hydrological basin differing in catchment dynamics, human impacts and limnological characteristics (Table 1) (Odada *et al.* 2003; Descy and Sarmiento 2008; Morana *et al.* 2014).

Sampling

Samples for limnological and biological parameters were collected during the rainy season in the pelagic zone of four lakes. Meromictic Lake Kivu (K) was sampled in February 2012 at two different sites: Ishungu (I) and the main basin of Gisenyi (G), located in the southern and northern part of the lake, respectively (Fig. 1, Table 1). Kivu samples were taken between 0 and 80 m depth, at intervals of 5 m (to cover the whole gradient of oxygen concentrations). Lakes Edward (E), Albert (A) and Victoria (V) were sampled in May 2012. They are shallower than Lake Kivu, holomictic and chlorophyll-*a* (Chl-*a*) concentrations are usually higher in their mixed layer (7.5 m, 20 m, 15 m; respectively) (Morana *et al.* 2014). Based on in situ observed vertical profiles of temperature and oxygen, the mixed-layer depth was determined in each water body. Water samples were collected every 5 m from the surface to the bottom of the mixed layer using a 7-L Niskin bottle (Hydro-Bios) and was then pooled to obtain a representative sample of this layer.

Limnological profiles were obtained in each lake using a Yellow Springs Instruments 6600 v2 multiparameter probe. Water samples for chemical and microbiological analyses were collected using a 7.5-L Niskin bottle and stored in 4-L plastic containers for chemical analyses (except for CH₄) and 2-L Nalgene plastic bottles for biological analyses. Filtrations were carried out in the field, filters and water samples were kept frozen until analyses in the laboratory.

Environmental variables

Water temperature (Temp), specific conductivity (Cond), depth (Z), pH and oxygen concentration (DO) were obtained from the limnological profiles. Total alkalinity (Alk) was carried out by open-cell titration with HCl 0.1 M on 50 mL water samples (Morana *et al.* 2014). Ammonia (NH₄⁺) concentrations were determined using the dichloroisocyanurate-salicylate-nitroprussiate colorimetric method. Nitrite (NO₂⁻) was calculated by the sulfanilamide coloration method, while

nitrate (NO_3^-) was determined after cadmium reduction to NO_2^- and quantified following the sulphanilamide coloration method (Llirós *et al.* 2010). NO_x was calculated as the sum of NO_2^- and NO_3^- , whereas DIN (dissolved inorganic nitrogen) was equivalent to $\text{NH}_4^+ + \text{NO}_2^- + \text{NO}_3^-$. Soluble reactive phosphorus (SRP) was determined using the molybdenum blue method (APHA 1992). Dissolved methane (CH_4) concentrations were measured by a headspace technique (Weiss 1981) using gas chromatography with flame ionization detection (GC-FID, SRI 8610C), as detailed by Borges *et al.* (2011). Chlorophyll-*a* (Chl-*a*) concentrations were determined by high-performance liquid chromatography (HPLC). Three liters of water were filtered on a Macherey-Nägel GF-5 filter (nominal porosity of 0.4 μm). Pigment extraction from the filters was carried out in 10 mL of 90 % HPLC grade acetone. After two sonication steps of 15 min separated by an overnight period at 4 °C, the extracts were stored in 2 mL amber vials at 25 °C. HPLC analysis was performed following the method described in Sarmiento *et al.* (2006) with a Waters system comprising a photodiode array and fluorescence detectors. Calibration was made using commercial external standards (DHI Denmark Lab Products). Precision for Chl-*a* measurement was better than $\pm 7\%$. Additional information on the analytical techniques and the data collected can be found in Morana *et al.* (2014).

DNA extraction, amplification of the 18S rRNA gene and pyrosequencing

Approximately 500 mL of surface mixed-layer water for DNA extraction from Edward, Albert and Victoria were prefiltered through 10 μm pore-size membranes (ISOPORE, Millipore, MA) and cells were collected onto 0.8 μm pore-size filters (ISOPORE, Millipore, MA). Moreover, based on the physical and chemical data, six depths of each site of Kivu (I and G) were chosen to analyze the diversity of microbial eukaryotes: I-5, I-15, I-30, I-45, I-50 and I-70 meters and G-5, G-15, G-25, G-30, G-35 and G-70 meters. In these cases, 1000 mL of water was passed through 5 μm pore size filters (ISOPORE, Millipore, MA) and planktonic cells were collected onto 0.2 μm pore-size membranes (ISOPORE, Millipore, MA). Each filter was placed in a cryovial contained 1.7 mL of lysis buffer (50mM Tris, 40mM EDTA and 0,75M Sucrose) and stored at -80 °C until DNA extraction.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DNA was extracted from filters following the phenol-chloroform-isoamyl alcohol protocol. In brief, filters were incubated with Lysozyme (1 mg mL⁻¹ final concentration) at 37 °C for 45 min while slightly shaken. In addition, Proteinase K (0.2 mg mL⁻¹) and sodium dodecyl sulfate (10 % final concentration) were added and tubes were incubated at 55 °C for 1 hour while slightly shaken. The resulting lysate underwent two steps of phenol–chloroform-isoamyl alcohol (25:24:1) extraction: the phenol mixture was spin 10 min at 12000 rpm. Then, one extract of equal volume of chloroform-isoamyl alcohol (24:1) was spin during 10 min at 12000 rpm. After the last centrifugation step, the aqueous phase was collected, concentrated in an Amicon Ultra unit (Millipore) and washed with sterile deionized water. The total DNA extract was quantified using a Nanodrop-1000 spectrophotometer (Thermo Scientific) and stored at -80°C until further analysis (Casamayor *et al.* 2002; Massana *et al.* 2004).

The microbial eukaryote diversity was assessed by high-throughput sequencing the 18S rRNA gene. The eukaryotic universal primers TAREuk454FWD1 (5'-CCAGCASCYGCGGTAATTCC-3') and TAREukREV3 (5'-ACTTTCGTTCTTGATYRA-3') (Stoeck *et al.* 2010) were used to amplify the hypervariable V4 region (~380 bp). The primers included 454-specific adaptors, and the forward primer also carried the barcode (Supplementary Table S1). The procedures followed for PCR amplifications were previously described in detail by Pernice *et al.* 2016. In brief, we added 3.5 µL (15 ng µL⁻¹) of genomic DNA to PCR tubes containing dNTPs (0.2 mM), 1µL of the primers R and F indicated above at 0.5 mM, and the Phusion High-Fidelity DNA Polymerase kit (Thermofisher) following manufacturer's recommendations in a final volume of 20 µL. Reactions were carried out in an automated thermocycler with the following cycle: initial denaturation step at 98°C for 3 min, followed by 30 cycles of denaturation at 98°C for 35 s., annealing at 51°C for 40 s., extension at 72°C for 45 s; and a final extension at 72°C for 10 min. Amplicons were checked in a 1.5% agarose gel for successful amplification. Due to low amount of DNA found, some samples (G-70, I-50, I-70 and E) were precipitated with ethanol-sodium acetate and resuspended in 13 µL of sterile water. PCRs were performed with 15 µL of PCR mix, 4 µL of the environmental DNA and 1 µL of Primer specific for 454-pyrosequencing. Triplicate amplicons were pooled, purified and quantified using the Qubit dsDNA broad-range (BR) assays (Invitrogen). About 10 ng µL⁻¹ (30 µL final volume) of PCR product

were sent to the Research and Testing Laboratory (Lubbock, TX, USA; <http://www.researchandtesting.com>) for sequencing using 454 GS FLX (Roche-454 Life Sciences, 454 Life Sciences, Branford, CT, USA) with Titanium chemistry.

Processing 454 sequences

Pyrosequencing data generated from the 454-sequencing runs were processed using QIIME (Caporaso *et al.* 2010). We obtained 145,236 sequences (pyrotags), which were de-multiplexed using the barcode identifier in the forward primer. Pyrotags were between 150–600 bp long and those which were selected had no more than 4 mismatches in the primer and no homopolymers longer than 8 bp. For quality check, errors were computed in sliding windows of 50 bp and pyrotags containing a window with an error >1% and appearing only once in the dataset were removed. After eliminating possible PCR and pyrosequencing errors a total of 117,432 reads from the 15 samples were left which were denoised with DeNoiser and clustered into OTUs (Operational Taxonomic Units) with UCLUST (99% similarity threshold) (Edgar 2010). OTUs were taxonomically assigned using PR2 (version 4.13.0) database (Guillou *et al.* 2013). OTUs were assigned to a given group when its representative sequence had a BLAST hit with an e-value below 10^{-100} against a reference sequence. Chimera check and removal was performed using the SILVA 108 reference database (Quast *et al.* 2013) within the MOTHUR program (Schloss 2008). The commands "align.seqs" and "chimera.slayer" were used (Haas *et al.* 2011) and Archaea, Bacteria and Metazoa sequences were eliminated. After removing singletons, doubletons, 1,003 OTUs with No Blast Hit (related to sequencing errors), and the sample G-30 (because it had less than < 1000 reads), the final out Table included 86,463 reads from 14 samples distributed among 779 OTUs.

Raw data was deposited at the European Nucleotide Archive public database under the following accession number: PRJEB42983.

Novelty analysis

To determine the degree of "novelty" of the OTUs, the sequences were aligned to the NCBI database using BLASTN, with default parameters. Then we obtained the similarity of the first best hit

1
2
3 253 corresponding to a cultured organism or to an environmental sequence, to assign the closest cultured
4
5 254 match (CCM) and the closest environmental match (CEM), respectively (Massana *et al.* 2011; Triadó-
6
7 255 Margarit and Casamayor 2012). After examining the averaged identity values for all sequences into
8
9 256 separate taxonomic groups, both similarity values (CCM and CEM) were plotted in a 2D dispersion
10
11 257 graph, giving the degree of “novelty” of the dataset. OTUs with high percentages of CEM indicate
12
13 258 that they are similar to sequences obtained in other environmental surveys, and those with low CEM
14
15 259 similarity highlight true novel diversity detected here. Likewise, points with high CCM similarity
16
17 260 indicate that sequences are close to cultured organism’s sequences, whereas dots with low CCM
18
19 261 similarity highlight environmental sequences with no cultured counterpart. Finally, for each plot we
20
21 262 defined “the highest novelty” as the area of the plot that contained OTUs with <90 or 95 %
22
23 263 (depending the supergroup) similarity to both CEM and CCM (del Campo and Massana 2011).
24
25
26 264

28 265 *Data Analysis*

30 266 The trophic status of the lakes was quantified by applying the Trophic State Index (TSI)
31
32 267 developed by Carlson (1977), using Chl-*a* as a variable. Richness was evaluated by rarefaction
33
34 268 analyses as the estimated number of OTUs in each lake (Hurlbert 1971). Rarefaction curves were
35
36 269 constructed using “rarecurve” function of Vegan version 2.5-6 (Oksanen *et al.* 2019) and R 3.5
37
38 270 (<http://cran.r-project.org>). Therefore, the relative abundance of each sample was re-sampled at the
39
40 271 minimum sample size (n = 1064 and 1628 at superficial level and Kivu depths, respectively) with the
41
42 272 “rrarefy” function, R package Vegan version 2.5-6 (Oksanen *et al.* 2019). The α -diversity (i.e.
43
44 273 Shannon_H and Simpson_1-D indices, as well as Pielou’s evenness) was determined using the Past
45
46 274 4.03 statistical package (Hammer 2016). Moreover, based on similar studies, we defined abundant
47
48 275 OTUs as those with a mean relative abundance in all samples $\geq 1\%$ (Logares *et al.* 2014; Du *et al.*
49
50 276 2019; Annenkova *et al.* 2020).
51
52

53 277 To explore the environmental drivers shaping protistan communities in surface mixed-layer
54
55 278 waters and Kivu depths, we performed multivariate analyses. Ordinations by detrended
56
57 279 correspondence analysis (DCA) indicated a linear distribution of the data (ter Braak and Šmilauer
58
59 280 1998). Then, a principal component analysis (PCA with supplementary variables) was conducted on
60

microbial communities interpreting this variation with the help of environmental variables, in order to add supplementary information on the scatter plot for a better understanding of the data (Lê *et al.* 2008). Supplementary variables were transformed with $y' = \ln(y + 1)$ in order to use linear ordination methods (ter Braak 1987). Variables characterized by high inflation factors were then dismissed. Since the microbiome data sets are compositional, they were not transformed and the PCA was the most accurate analysis representing the relationships between the eukaryotic microorganisms and the distances between samples on a common plot (Gloor *et al.* 2017). Multivariate analysis was performed using the CANOCO 5 software (ter Braak and Šmilauer 1998).

Phylogenetic exploration of "novel" diversity in Lake Kivu

According to "Novelty analysis", Alveolata and Stramenopiles were two of the supergroups that showed the highest numbers of novel OTUs. Thereby, OTUs belonging to those lineages in Lake Kivu were classified phylogenetically. First, we performed a reference alignment using the sequences from Mahé *et al.* (2017) reference tree and the best hit from each OTU obtained by the taxonomic assignment explained before. The alignment was performed using MAFFT version 7.453 (Katoh *et al.* 2002; Katoh and Standley 2013) with the E-INS-i algorithm and revised with Geneious version 9.0.5 (www.geneious.com). After obtaining a reference alignment, the OTUs sequences were added with the function --addfragments from MAFFT. The final alignment was trimmed automatically using TrimAl version 1.4.rev22 (Capella-Gutiérrez *et al.* 2009) with -gt 0.3 and -st 0.001 parameters. We inferred the maximum-likelihood tree using IQ-Tree version 1.6.12 (Nguyen *et al.* 2015) with 1000 bootstrap replicates for SH-aLRT. The best model selected automatically by IQ-Tree was GTR+F+R10 (Kalyaanamoorthy *et al.* 2017; Nguyen *et al.* 2015). The final phylogenetic tree was visualized and annotated using Itol (Letunic and Bork 2019) and inkscape (<https://inkscape.org/>).

Results

Environmental parameters

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The East African Great Lakes displayed some remarkable differences in lake area, water depth, physical, chemical and biological parameters (Table 1 and 2).

a) Surface mixed-layer waters

During the sampled period, the trophic status of these water bodies varied from meso-eutrophic systems (Lake Edward) to less productive waters (Lake Kivu). Water temperatures were typical of tropical lakes (24.1°C at Ishungu-27.9°C in Lake Albert), whereas DO values ranged from 5.7 mg L⁻¹ in Ishungu to 8.5 mg L⁻¹ in L. Albert. Moreover, a wide range of conductivity values was observed, L. Kivu being up to two orders of magnitude higher than the other lakes (95 µS cm⁻¹ at Victoria-1155 µS cm⁻¹ at Ishungu) (Table 2).

b) Vertical profile in Lake Kivu

Lake Kivu profiles were characterized by a vertical gradient of several physical and chemical parameters, some of which differ between basins (Fig. 2). In general, surface waters displayed high DO but low nutrients concentrations; while CH₄ and conductivity grew gradually with depth, whereas pH decreased over the first 80-m depth, more drastically in Ishungu. The low mixing regime during the rainy season allowed the establishment of the epilimnion, with a mixed layer depth which varied between 15 m in Ishungu to 25 m depth in Gisenyi (the main basin). This basin showed a steep oxycline from 25 to 40 m depth (transition zone), whereas the DO concentration profiles in Ishungu decreased more smoothly, with an oxycline extending from 15 to 50 m depth. Anoxic waters were detected from 45 and 55 m, in Gisenyi and Ishungu, respectively (Fig. 2 A,B). NO_x (NO₂⁻ + NO₃⁻) concentrations were low throughout most of the water column for both basins, but an accumulation occurred in the oxic–anoxic transition zone with a peak at 35 m in Gisenyi (2.3 µM) and 45 m in Ishungu (9.6 µM) (Fig. 2 C,D).

Overall composition of protist communities

Across the 14 samples, we obtained a total of 86,463 quality-filtered sequences, derived from a variable sequencing effort among samples that ranged from 1064 sequences in Lake Victoria to 15,105 Ishungu at 15 m. These reads were grouped in 779 OTUs at 99% sequence identity. Rarefaction curves of OTU richness for each lake and depth indicated that some samples did not reach saturation, suggesting insufficient sequencing depth to capture all protist diversity (Supplementary Fig. S1A,B).

Total reads were spread among eleven high-rank supergroups or protistan lineages and grouped in 57 eukaryotic groups or classes. Most of the sequences were affiliated to Alveolata (31% OTUs, 44% reads), Opisthokonta (20% OTUs, 8% reads) and Stramenopiles (17% OTUs, 10% reads), while unclassified eukaryotes (Eukaryota) (3% OTUs, 0.6% reads), Amoebozoa (1% OTUs, 0.5% reads) and Telonema (0.3% OTUs, 0.09%) were the least represented (Fig. 1B). In terms of numbers of groups, the most diverse was Stramenopiles (15 groups), with Chrysophyceae displaying the highest percentage of OTUs (45%), followed by Alveolata (9 groups), dominated by Ciliophora (54% of OTUs) and Opisthokonta (8 groups), dominated by Basal_Fungi (47%), which comprises a diverse group of heterotrophic, saprophytic and parasitic organisms such as Microsporidiomycota and Chytridiomycota (Supplementary Fig. S2).

Potential "novel" diversity

Approximately 40.6% of OTUs (15.5% reads) displayed low CCM and CEM similarity (< 90 or < 95%, depending on the supergroup) with reported sequences in public databases (Fig. 3). Alveolata (38% OTUs, 35% reads), Opisthokonta (16% OTUs, 15% reads) and Stramenopiles (10% OTUs, 15% reads) showed the highest "novel" diversity (Fig. 1B). Among alveolates, 29% of OTUs affiliated more closely with Ciliophora, within Opisthokonta, 18.7% of OTUs were associated with Choanomonada, whereas among Stramenopiles, 13% of OTUs were more closely affiliated to Chrysophyceae.

Protists communities in surface mixed-layer waters

1
2
3 361 The α -diversity indices differed in surface mixed-layer of East African Great Lakes (Table 3).
4
5 362 L. Victoria displayed the most dissimilar values. Protists OTUs richness ranging from 96 in L. Kivu
6
7 363 off Gisenyi to 125 in L. Victoria. Similarly, the main basin of L. Kivu had the lowest and highest
8
9 364 diversity values, respectively (inverse Simpson index: 0.85-0.93, Shannon index: 2.97-3.59;
10
11 365 respectively). While the evenness ranged from 0.17 in L. Edward to 0.29 in L. Victoria.

13
14 366 The number of reads varied widely, with the lowest number in L. Victoria (1064) and the
15
16 367 highest in L. Edward (13,129) (Supplementary Fig. S1A). In order to compare the microbial
17
18 368 eukaryote composition among all lakes, we standardized the samples at the smallest number. The
19
20 369 standardized dataset resulted in an overall number of 5,320 reads. The relative numbers of reads of
21
22 370 taxonomic groups varied between lakes; however, reads affiliated to Alveolata (mean 40.4%),
23
24 371 Ciliophora (mean 23%) and Dinoflagellata (mean 17.7%) being the most abundant, followed by
25
26 372 Cryptista (mean 30.8%), mainly Cryptomonadales (mean 26.2%), accounted for the majority of
27
28 373 sequences in all lakes (Fig. 4A). Amoebozoa being found in I-5 (0.28%) and V (0.09%), while only 2
29
30 374 OTUs more closely affiliated to Telonema were detected only in E (0.17%). Those were the least
31
32 375 abundant supergroup (not displayed in the figure). On average, 12.7% of reads were considered
33
34 376 "novel" in the surface mixed-layer waters.

36
37 377 Considering the most abundant OTUs (OTUs $\geq 1\%$), only 14 (66.7% of total reads) were
38
39 378 dominant (Supplementary Fig. S3A). Lakes Albert and Victoria, following by L. Edward had a
40
41 379 similar composition in the proportion of eukaryotes, whereas both basins of L. Kivu differed in their
42
43 380 protistan assemblage. Ishungu was characterized by the dominance of OTU_620, affiliated closely to
44
45 381 Cryptomonadales, while Gisenyi for the occurrence of OTU_1404 (Dinoflagellata). The potential
46
47 382 "novel" OTU_1814 affiliated to Cryptomonadales with 86.7% of similarity (representing 6.6% of the
48
49 383 total reads) was observed in different proportions in all lakes.

51
52 384 Multivariate analysis conducted on microbial eukaryote communities showed a segregation of
53
54 385 lakes at the surface level (Fig. 4B). The Principal Component Analysis (PCA) plot over the first two
55
56 386 axes (70.5% of total variation) revealed L. Kivu as the most dissimilar, with the two sites separated
57
58 387 from the rest of the water bodies. The first axis explained 40.55 % of total cumulative variation and it
59
60 388 was mainly associated positively with altitude ($r = 0.87$), maximum depth ($r = 0.71$) and conductivity

($r = 0.59$), and negatively with Chl-*a* ($r = -0.94$), DIN ($r = -0.84$), temperature ($r = -0.83$), DO ($r = -0.63$) and SRP ($r = -0.44$). The second axis (29.93 % of total variance) was related mainly by Cond ($r = 0.48$). The PCA showed a positive association between the variables SRP, DIN and Chl-*a* and the lakes Albert and Victoria, suggesting productive systems. These lakes displayed the most similar community structure distinguished by the presence of Diplonema and MALVs (marine alveolates), as well as a large number of Chrysophyceae and Kinetoplastida; among others. Diversity in L. Edward was mainly explained by Chl-*a* and showed a diverse microbial eukaryote community. On the other hand, L. Kivu was associated with higher values of conductivity, depth and altitude. In particular, at Ishungu a more heterotrophic community was found, including fungi and parasites, while Gisenyi displayed a higher abundance of MASTs (marine Stramenopiles), Charophytes and Dinoflagellates. Further, only 18 OTUs were shared by all the freshwaters systems, most of them being abundant (61.1% of reads) (Fig. 4C), with ciliates and cryptomonadales as the most represented (27.7% of the total). In fact, approximately 64% of the OTUs observed in the surface mixed-layer waters was specific to a single lake.

Vertical profiles of Lake Kivu's protists communities

In L. Kivu, the normalized dataset resulted in an overall number of 17,908 reads in 11 separate samples, corresponding to 544 OTUs. The southern part of the lake (Ishungu) was more diverse (Table 4), with higher values of α -diversity indices at 70 m (Richness: 216, inverse Simpson index: 0.98, Shannon index: 4.45, Evenness: 0.40).

Most OTUs affiliated to Alveolata (mean 47.6%) and Cryptista (mean 22.6%). Along the first 30 m of Ishungu, Cryptista (40.3% of total reads) dominated, being Cryptomonadales (38.7% of total reads) the most abundant group (Fig. 5A). After 45 m depth, there was a change to a more heterotrophic protist community. At 45 m (I-45), 38.9% of reads corresponded to Dinoflagellata (Alveolata) and 36.9% affiliated to diverse heterotrophic groups; whereas in I-50 and I-70 Chrysophyceae (Stramenopiles) appeared in greater proportion (18.3%). Conversely, in Gisenyi the abundance of Alveolata gradually declined over the vertical profile, being Dinoflagellata (mean

1
2
3 416 26.4%) and Ciliophora (mean 18.3%) the most important groups (Fig. 5B). The potential microbial
4
5 417 eukaryotic novelties were highest in deeper samples at both sites (33.4% on average at 70 m).
6

7 418 Multivariate analysis revealed a great dissimilarity between sites and depths (Fig. 5C). The
8
9 419 PCA plot with supplementary variables accounted for 76.2% of the total variance. Axis 1 explained
10
11 420 45.7% and was positively related to Temp ($r = 0.30$) and negatively to Chl-*a* ($r = -0.38$). Axis 2
12
13 421 explained 36.3% and was positively related to Cond ($r = 0.58$) and negatively to DO ($r = -0.64$) and
14
15 422 Temp ($r = -0.46$). Microbial eukaryote communities were distributed according to physical-chemical
16
17 423 parameters. The shallower samples (up to 35 meters) corresponding to the two sites were distributed
18
19 424 in the lower part of the diagram mainly associated with higher temperatures, DO and Chl-*a*
20
21 425 concentrations. The deepest samples from Ishungu (I-50 and I-70) forming a separate group in the
22
23 426 upper left side of the plot mainly related to high conductivity values; while I-45 were the most
24
25 427 isolated, more associated with high nitrate concentrations, presenting a particular assemblage of
26
27 428 microorganisms, many of which were heterotrophic.
28
29

30 429 Taking into account the composition of most abundant OTUs (OTUs $\geq 1\%$), only 24 were
31
32 430 dominant (62.5% of total reads), which also differed between sites and depths (Supplementary Fig.
33
34 431 S3B,C). Usually, Ishungu displayed a more dissimilar eukaryotic microbial community through
35
36 432 depths. The firsts 30 m of this basin were dominated by the OTU_620 (Cryptomonadales); whereas
37
38 433 OTU_1404 (Dinoflagellata) abounded in Gisenyi. Besides, three "novel" OTUs closely affiliated to
39
40 434 Ciliates were also noteworthy (OTU_452, OTU_394 and OTU_2540).
41
42

43 435 The "novel" OTUs affiliating to Alveolates and Stramenopiles in L. Kivu were explored by
44
45 436 phylogeny. In total 327 sequences were retrieved from NCBI and aligned to construct two reference
46
47 437 alignments. OTUs affiliated to Alveolata (201 OTUs) and Stramenopiles (98 OTUs) were added to
48
49 438 the respective reference alignment for their phylogenetic classification. Most Stramenopiles OTUs
50
51 439 were affiliated to Chrysophyceae (14 OTUs) and Bicosoecida (9 OTUs), mostly observed at great
52
53 440 depths in both basins of L. Kivu (I-70 and G-70) (Fig. 6). The "novel" OTUs 136, 917, 1941 and
54
55 441 2316, together with sequences from deep anoxic marine environments (HM749941 and AB505560)
56
57 442 were grouped in a well-supported environmental cluster mostly observed in G-70. "Novel"
58
59 443 Chrysophyceae OTUs were related to Clades C, F2, F1 and A, formed mostly by heterotrophic
60

organisms from freshwater and marine environments. Furthermore, 6 "novel" OTUs were affiliated to MASTs, 3 of which were grouped into MASTs 12 (OTU_1894 and 1644 to MAST-12C, OTU_1766 to MAST-12A), 2 OTUs were related to MAST 4/7 (OTU_1606 to MAST-7B and OTU_693 to MAST-4A), and OTU_1445 was affiliated to MAST-3J together with EU561853.2, a sequence from uncultured marine eukaryote.

Among Alveolata (Fig. 7), 87 OTUs were classified as "novel", most of them related to Ciliophora (56 OTUs) and Dinoflagellata (19 OTUs). Dinoflagellate "novel" OTUs were mostly found at surface level (at 5 and 15 m), while most Ciliophora "novel" OTUs were observed in deeper samples (I-50, I-70 and G-70), highly supported by clades closer to Prostomatea (Fig. 8). Moreover, a great diversity of "novel" OTUs was related to Oligohymenophorea and 15 OTUs affiliated to MALVs (Syndiniales). Among them, 4 to MALV-I, 10 to MALV-II and 1 to MALV-III, (Supplementary Table S2); most of them were observed with relative high abundance at I-70 and G-70 samples (Fig. 7).

Discussion

Despite the key role played by protists in aquatic ecosystems (Sherr and Sherr 1988), their diversity and function is still underestimated in freshwater (Schloss *et al.* 2016; Debroas *et al.* 2017), even more in meromictic lakes (e.g Tarbe *et al.* 2011a; Lepère *et al.* 2016), although they provide a high diversity of niches along stratification gradients (Oikonomou *et al.* 2014). East-African Lakes are extremely old and their large biodiversity is supported by complex processes that include potential endemic organisms (*i.e* Seehausen 2006; Descy and Sarmiento 2008). Our study revealed a high and potential "novel" diversity of small eukaryotes ($< 10 \mu\text{m}$) in African Great Lakes, associated with different physicochemical variables, especially in the meromictic Lake Kivu.

Limnological features of these lakes are similar to those previously reported (Pasche *et al.* 2009; Sarmiento *et al.* 2008; Morana *et al.* 2014). Although all systems are tropical, they differed concerning abiotic parameters, catchment dynamics, human impacts and food webs (Odada *et al.* 2003).

1
2
3 471 Protists communities are poorly known in African freshwaters (Llirós *et al.* 2012). A high
4
5 472 diversity was found in an earlier study by Tarbe (2010) in surface waters of Tanganyika and Kivu.
6
7 473 Accordingly, a large number of protistan lineages or supergroups were revealed in our work. Over the
8
9 474 past decades, the eukaryote Tree of Life has been divided into several supergroups and multiple
10
11 475 lineages of uncertain placement (Massana *et al.* 2011; Ortiz-Álvarez *et al.* 2018; Burki *et al.* 2021).
12
13 476 The configuration of supergroups varies constantly as more studies are conducted. Recently, Burki *et*
14
15 477 *al.* (2021) rearranged the "tree" and the term Hacrobia, previously described by Okamoto *et al.*
16
17 478 (2009), was split into Haptista and Cryptista, while Telonema became part of TSAR. Phylogenomic
18
19 479 analyses infer a Metamonada plus Discoba clade corresponding to Excavata, and some ones recover
20
21 480 instead a specific relationship with the excavate group malawimonads. As these relationships are still
22
23 481 uncertain, we describe our results using the original term Excavata and we incorporate these new
24
25 482 findings as the eukaryotic phylogeny progressively resolves (Fig. 1B). Thereby, in African lakes,
26
27 483 assemblages were dominated by Alveolata, Opisthokonta and Stramenopiles, according to Tarbe
28
29 484 (2010) and those reported from other freshwater systems (Triadó-Margarit and Casamayor 2012;
30
31 485 Simon *et al.* 2015; Ortiz-Álvarez *et al.* 2018). A recent study revealed that phototrophs dominated in
32
33 486 freshwaters, while parasitic taxa represented roughly 15 to 20% of all sequences in marine and soil
34
35 487 ecosystems (Singer *et al.* 2021). Here, a large number of heterotrophic phagotrophs and parasites
36
37 488 organisms were also found and our results more closely resemble those found in Lake Baikal, the
38
39 489 oldest lake in the world, as we will discuss in detail below (Annenkova *et al.* 2020; David *et al.* 2021).
40
41
42
43 490

45 491 *Potential "novel" diversity*

47 492 A high percentage of potential "novel" organisms was found, since 40.6% of OTUs had low
48
49 493 similarity with reported sequences in public databases (Triadó-Margarit and Casamayor 2012;
50
51 494 Massana *et al.* 2015). Previous studies in some groups have already shown that "novelty" is larger in
52
53 495 freshwater than in marine systems (del Campo and Massana 2011), probably due its heterogeneous
54
55 496 abiotic parameters compared to the relatively buffered conditions in the ocean (Singer *et al.* 2021). In
56
57 497 high mountain lakes, a high "novel" protists, mainly for Rhizaria and Opisthokonta supergroups, as
58
59 498 well as Alveolata and Stramenopiles was reported (Triadó-Margarit and Casamayor 2012; Ortiz-

Alvárez *et al.* 2018). A high degree of "novelty" has been also found in picoeukaryotes from pampean lakes in Argentina (Metz *et al.* 2019). The percentage found in our work is a slightly greater than that recorded, likely explained by the fact that freshwater tropical environments are undersampled (Llirós *et al.* 2012; Sarmiento 2012). Likewise, some groups could be endemic, since in ancient lakes microorganisms could have had time to diversify, as already reported for other species in these lakes (Seehausen 2006) and for protists in Baikal (Annenkova *et al.* 2020). The environmental heterogeneity and persistently extreme conditions might promote an adapted specialist microbiota (Catalan *et al.* 2006). Regardless, all these studies highlight that freshwater planktonic diversity remains largely to be explored.

Protists community in surface mixed-layer waters

A high microbial diversity and richness in surface mixed-layer waters was found, probably to a wide variation in the physical and chemical properties (Debroas *et al.* 2017; Boenigk *et al.* 2018). The eukaryote assemblage was dominated by Ciliophora and Dinoflagellata, followed by Cryptomonadales (Fig. 4A, Supplementary Fig. S3A), as reported in tropical and freshwaters systems (Tarbe *et al.* 2011b; Kammerlander *et al.* 2015; Simon *et al.* 2016; Debroas *et al.* 2017). Usually, alveolate reads are found in smaller fractions because they have a high copy number of the SSU rRNA gene, that presumably break down and pass through the filter (Dyal *et al.* 1995; Galluzi *et al.* 2004). Nevertheless, small organisms were also observed under the microscope and the relative abundance of groups varied between lakes.

The mesotrophic lakes Albert and Victoria presented a similar eukaryote community structure. Among Discoba (Excavates), Diplonemea and Kinetoplastida were distinguished. These two heterotrophic flagellates are generally abundant in the sea, but were recently reported in freshwaters (Boenigk *et al.* 2015; Pernice *et al.* 2016; Mukherjee *et al.* 2019). In particular, some groups previously thought to be exclusively marine have been identified in the plankton of Lake Baikal (David *et al.* 2021). In general, Excavata are very poorly represented when the V4 region is amplified (Obiol *et al.* 2020), therefore our abundances are likely to be underestimated. Moreover, MALV-I, -II and -III, inhabiting frequently marine environments (Massana *et al.* 2011; de Vargas *et*

1
2
3 527 *al.* 2015), were also found. Those groups are extremely rare in freshwaters. However, MALVs were
4
5 528 recorded in a meromictic Arctic lake (Charvet *et al.* 2012) as well as in Baikal (Annenkova *et al.*
6
7 529 2020; David *et al.* 2021). The potentially "novel" OTU_1814, closely affiliated to Cryptomonadales
8
9 530 (86.7% of similarity), was rather abundant in L. Victoria (Supplementary Fig. S3A). On the other
10
11 531 hand, L. Edward was associated with higher Chl-*a*, temperature, DO and nutrient values in the PCA
12
13 532 (Fig. 4B). These characteristics support the growth of a large number of microorganisms including
14
15 533 aerobic and heterotrophic protists, like Ciliophora and Cercozoa (Liu *et al.* 2021). In this lake, within
16
17 534 minority groups also highlights the small generalist Katablepharidae, which is usually found in low
18
19 535 abundances in lakes and oceans (Simon *et al.* 2015; Boenigk *et al.* 2018) and Dictyochophyceae,
20
21 536 almost exclusively marine photosynthetic algae, also found in lakes Taihu (Chen *et al.* 2008) and
22
23 537 Baikal (Annenkova *et al.* 2020). These organisms possibly belong to Pedinellales, common in fresh
24
25 538 waters (Liu *et al.* 2021).

26
27
28 539 Both basin of Lake Kivu exhibited a dissimilar community structure. Associated with high
29
30 540 conductivity, scarce nutrients, low temperature and Chl-*a* concentrations, the oligotrophic L. Kivu
31
32 541 harbors unique limnological conditions and particular protistan assemblages. Surface waters are
33
34 542 usually colder than the rest of the Rift lakes, as a consequence of its elevation (1463 m) (Schmid *et al.*
35
36 543 2005). Ishungu is located in the smaller southern basin which is less exposed to winds and harbors a
37
38 544 large amount of Cryptomonadales. These mixotrophic flagellates can survive under low light
39
40 545 condition, allowing the exploitation of oligotrophic environments (Oikonomou *et al.* 2014; Boenigk *et*
41
42 546 *al.* 2015). Here, the percentage of alveolates decreased to 24.5% and Perkinsidae (0.10%) were
43
44 547 detected. Opisthokonta reached 9.6%, being Ascomycota, Basal_Fungi, Ichthyosporea and
45
46 548 Syssomonas the most abundant groups. A large number of fungi were observed, in contrast to
47
48 549 temperate lakes (Lefèvre *et al.* 2007; Lepère *et al.* 2008) and the findings by Tarbe *et al.* (2011b) in
49
50 550 Tanganyika. Fungi is one of the most understudied microbial groups in aquatic systems, therefore its
51
52 551 distribution and role in freshwaters environments are largely unknown. Recently, Lepère *et al.* (2019)
53
54 552 found a high diversity in lakes and rivers, suggesting that freshwater fungi have been undersampled.
55
56 553 In addition, unlike Tanganyika (Tarbe *et al.* 2011b), Bicosoecida were also registered. Ishungu is
57
58 554 located in a bay close to the only outflow through the Ruzizi River that discharges into Lake
59
60

1
2
3 555 Tanganyika (Sarmiento *et al.* 2012). Its proximity and the same trophic state between both lakes could
4
5 556 suggest a similar protist assembly. Nevertheless, as food webs are different, small eukaryotes
6
7 557 communities seem to be specific, as found Tarbe (2010) with less than 11% of shared sequences. The
8
9 558 thermal stratification of the water column reduces nutrient supply in the euphotic zone, which is
10
11 559 generally shallower at Ishungu, limiting phytoplankton growth (Descy *et al.* 2012; Sarmiento *et al.*
12
13 560 2012). All these characteristics could favor mixotrophy, phagotrophy and parasitism as a strategy of
14
15 561 organisms to obtain energy in this site.

16
17
18 562 The deepest basin, Gisenyi, presented a dissimilar microbial eukaryotic communities
19
20 563 compared with Ishungu. Conductivity was higher in this site (Schmid *et al.* 2005; Olapade and
21
22 564 Omitoyin 2012), favoring thecate organisms. Among most abundant taxa, the OTU_2553 was related
23
24 565 to a dinoflagellate (99.5% of similarity), already found in Tanganyika (Tarbe *et al.* 2011b).
25
26 566 Dinoflagellates distribution is known to be influenced by light, temperature and salinity, being
27
28 567 commonly distributed in the sea as well as in large-freshwater bodies (Tarbe *et al.* 2011a; Khomich *et*
29
30 568 *al.* 2017; Órtiz-Alvarez *et al.* 2018). Among less frequent groups, two Archaeplastida (Chlorophyceae
31
32 569 and Charophyceae) and two Stramenopiles (Diatomea and MASTs) were detected. Chlorophyceae are
33
34 570 a very important group of algae in continental aquatic systems (Khomich *et al.* 2017; Metz *et al.*
35
36 571 2019), but in terms of abundance, they tend to be poorly represented in most African great lakes
37
38 572 (Tarbe *et al.* 2011b). Charophyta are structurally complex green algae, frequently found in hard
39
40 573 waters, like Baltic Sea (Schubert and Blindow 2004) and in tropical brackish lagoons impacted by
41
42 574 humans (Palma-Silva *et al.* 2004). Diatoms are among the major groups of photosynthetic organisms,
43
44 575 contributing to a large part of the Earth's primary production, inhabit a wide variety of environments
45
46 576 (Verleyen *et al.* 2009) and already observed by Sarmiento *et al.* (2006). On the other hand, we found a
47
48 577 few OTUs belonging to 'typically marine' lineages MAST-3 and MAST-4 (Massana *et al.* 2004;
49
50 578 Logares *et al.* 2012), although some of them have already been registered in fresh waters (Charvet *et*
51
52 579 *al.* 2012; Simon *et al.* 2014; Arroyo *et al.* 2018).

53
54
55 580 Overall, the lakes showed a small eukaryote community structure associated to conductivity
56
57 581 values and productivity gradients. A similar assemblage associated with the most productive lakes
58
59 582 Albert, Victoria and Edward was found in temperate lakes (Lefranc *et al.* 2005; Simon *et al.* 2015).

On the other hand, differences between sites in the oligotrophic L. Kivu seemed to be due to OTUs $\geq$ 1%, which generally contribute to biomass production, as well as minors groups, that could regulate the functioning of aquatic habitats (Debroas *et al.* 2015; Du *et al.* 2019). Indeed, Ishungu presented a more heterotrophic community, whereas the microbial food web would be supported by autochthonous organic carbon sources in Gisenyi.

How is the community structure along the depth gradient in both basins of Lake Kivu?

The water column structure in L. Kivu is different from most other large lakes in the world (Schmid and Wüest 2012). Because of their morphometry, the influence of the nearby volcanoes, groundwater inflows and anthropogenic activities, both basins differed in vertical abiotic parameters (Fig. 2) (Llirós *et al.* 2012; İnceoğlu *et al.* 2015), providing different conditions for life by microbial eukaryote communities down the water column. These conditions were reflected in our study, with noticeable differences between sites and depth, as demonstrated in an Arctic meromictic lake (Charvet *et al.* 2012) and in Lake Baikal (Annenkova *et al.* 2020).

The first 30 m of Ishungu were dominated by Cryptomonadales ($\approx$ 40% of the total reads), a generalist taxa occurring in many environmental conditions (Boenigk *et al.* 2018). Moreover, we highlight the presence of Kinetoplastida, being OTU_1531 among the most abundant (Supplementary Fig. S3B). This group is ecologically important because of their diverse mode of feeding (Salani *et al.* 2012), containing many bacterivorous (e.g. *Bodo*) as well as endoparasitic lineages (e.g. *Trypanosoma*), who could infect fishes, with the consequent economic losses (Corrêa *et al.* 2016). Associated with different habitats within marine and freshwater environments (Yubuki and Leander 2018), Kinetoplastida were previously reported by Tarbe *et al.* (2011a) in the epilimnion of Tanganyika and Kivu, and in our study *Trypanosoma* spp. was detected with 85% of CCM similarity. From 45 m depth down, microorganism's diversity changes abruptly to a more heterotrophic community (Table 4, Fig. 5A). At 45 m depth, a NO_x peak was recorded (Fig. 2C) and the oxic–anoxic transition zone was dominated by the nitrite-oxidizing *Nitrospira* (İnceoğlu *et al.* 2015). Here, dinoflagellates (38.8% of total reads) dominated, being OTU_1404 the most abundant (Supplementary Fig. S3B). Below this zone (I-50 and I-70), anoxic waters prevailed. Others studies

have shown that a diverse methanotrophic bacterial community was highly active in these layers (İnceoğlu *et al.* 2015; Morana *et al.* 2016), which could support a complex assemblage of protists (Taylor *et al.* 2006; Llíros *et al.* 2012). Indeed, a greater diversity was found, dominated by heterotrophic and mixotrophic organisms (Ciliates and Chrysophyceae), saprotrophic (Fungi), parasitic (Perkinzoa) and many other minor heterotrophic groups.

In the epilimnion of Gisenyi, the community structure was dominated by dinoflagellates and ciliates. The decline of these groups at 25 m could have been due to an increase of Perkinsidae, which reached 14% of total reads (Fig. 5B). Perkinsidae was currently found in marine systems (Massana *et al.* 2015), but many of them also began to be reported in fresh waters (Lefranc *et al.* 2005; Ortiz-Alvarez *et al.* 2018), and also in the water column of L. Baikal (David *et al.* 2021). This parasitic group can infect a wide variety of organisms. Parasites infecting protists as ciliates and dinoflagellates could directly influence the food web (Lepère *et al.* 2008; Ortiz-Alvarez *et al.* 2018; Cruaud *et al.* 2019). Thus, the high relative abundance found specifically at this depth could imply the importance of parasitism in regulating these protists. As in Ishungu, a high percentage of "novel" OTUS were found in the anoxic waters.

Overall, the protistan assemblages in both basins in Lake Kivu were dissimilar, reflecting differences in abiotic parameters along the depth profile. Studies in tropical and meromictic lakes are scarce, therefore this work provides new insights into these specific ecosystems. We revealed a high genetic diversity, particularly in anoxic waters, as was already reported in marine (Pernice *et al.* 2016; Zhao *et al.* 2017) and in some meromictic systems (Oikonomou *et al.* 2014; Lepère *et al.* 2016). Photosynthetic algae can inhabit extreme environments, being able to survive as cysts (Lepère *et al.* 2016). However, the community of the lower mixolimnion layers was composed mainly by heterotrophic, mixotrophic and saprotrophic organisms, highlighting their important role in the food web (Lepère *et al.* 2008). Furthermore, the extremely high potential "novelty" of taxa recorded in the deeper layers was remarkable, particularly in Alveolata and Stramenopiles supergroups. These taxa also showed a substantial proportion of clones (10–20%. c. 162 sequences) with low relatedness with any previously reported sequence in GenBank in Pyrenean lakes (Triadó-Casamayor *et al.* 2012). Also, in a meromictic lake, unclassified ciliate had highest representation at 29 m (Charvet *et al.*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

2012). Our results seem to be more closely similar to those also found in L. Baikal, where also depth, and consequently nutrients, light, temperature and DO had a strong effect on the protistan assemblage, large size and longevity allowing *in situ* evolution (Annenkova *et al.* 2020; David *et al.* 2021).

Phylogenetic exploration of "novel" diversity in Lake Kivu

Lake Kivu hosted a great phylogenetic diversity with 280 OTUs classified as “novel”. Chrysophyceae OTUs were related to clades formed mostly by heterotrophic organisms from freshwater and marine environments (del Campo and Massana 2011). Bicosoecida, a common group in freshwater, was also found in hypersaline systems (Debroas *et al.* 2017; Filker *et al.* 2017) (Fig. 6). Most of them were related to freshwater reference sequences and distributed at different depths. OTU_69 and OTU_1968 were distantly related to *Cafeteria*, a common marine genus; however, they present a low similarity to their reference sequence (< 90%). Moreover, "novel" Ciliophora OTUs were many and diverse at great depths (I-50, I-70 and G-70) (Fig. 8). Ciliophora is an ubiquitous, complex and highly diverse group, already observed to display high "novelty" in other extreme environments (Reboul *et al.* 2018). "Novel" OTUs closer to Prostomatea affiliated to sequences from lake sediment (KP059214) and anoxic waters (GU819524). "Novel" OTUs related to Oligohymenophorea were mostly obtained from anoxic deep marine samples and three "novel" OTUs were classified as abundant ($\geq 1\%$): Oligohymenophorea (OTU_452), Phyllopharyngea (OTU_394) and an environmental OTU (OTU_2540). OTU_452 presented a relatively high contribution in I-45 and OTU_2540 in I-70 and G-70 samples, highlighting their importance in these environments and how little we know about the diversity of microbial eukaryotes in anoxic freshwater systems.

Marine-freshwater transitions

A large number of protist lineages, mainly related to marine environment, like MASTs, Perkinsidae, marine Ciliates, Chlorarachniophyta and MALVs among others, was found. With the advance of molecular techniques, MASTs are already being detected in different freshwater systems (Charvet *et al.* 2012; Simon *et al.* 2014; Arroyo *et al.* 2018). On the other hand, successful transitions from marine to fresh waters in Perkinsidae have been reported in detail (Bråte *et al.* 2010). This group

also began to be reported from high-mountain oligotrophic lakes (Ortiz-Álvarez *et al.* 2018) and recently a large distribution across freshwaters was observed, suggesting potential parasitic associations with phytoplankton (Jobard *et al.* 2020) and emphasizing the need of further studies of these parasitic organisms. Most "novel" ciliates were abundant in L. Kivu deep samples and they mostly affiliated to marine sequences. Those anoxic deep waters were characterized by high nutrients concentrations and conductivity (Pasche *et al.* 2009; Roland *et al.* 2017), harboring diverse and complex bacterial and archaeal assemblages (Llirós *et al.* 2010; İnceoğlu *et al.* 2015). The high density of ciliates in deep waters could be explained by the fact that many of them could prey on prokaryotes (Fenchel and Finlay 2010, Charvet *et al.* 2012). Likewise, anaerobic ciliates have been able to adapt to anoxic conditions by harboring methanogenic symbionts (Oikonomou *et al.* 2014). Another exclusive marine algae found is chlorarachniophytes, widely distributed in tropical and temperate waters (Ishida *et al.* 2007; Gile *et al.* 2010) and, until now, not registered in other freshwater habitats. Members of MALVs were registered in L. Baikal and in association with MASTs and other Cercozoa, they are likely to participate in complex interactions (Genitsaris *et al.* 2020). Hence, our results demonstrate a complex eukaryotic diversity, in which several lineages have adapted to completely different niches and have succeeded to overcome the salinity boundary (Annenkova *et al.* 2015; Simon *et al.* 2014; Arroyo *et al.* 2018).

Funding

This study was supported by the Spanish Ministry for Science and Technology (AGLOM project CGL2010-11556-E), Belgian Federal Science Policy Office – BELSPO (Project EAGLES), Fonds de la Recherche Scientifique – FNRS (Project MICKI - Microbial diversity and processes in Lake Kivu), 'Damas' mission, by FNRS, and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP process 2014/14139-3). PF was supported by the fellow "Bourse de séjour scientifique", by FNRS and Programa de Financiamiento Parcial de Estadías Breves en el Exterior para Becarios Postdoctorales by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET - Argentina). HS was supported by Conselho Nacional de Desenvolvimento e Pesquisa Tecnológica (CNPq productivity grant 309514/2017-7).

1
2
3 695 **Acknowledgments**

4
5 696 We thank Pascal Isumbisho (Institut Supérieur Pédagogique, Bukavu, Democratic Republic
6
7 697 of Congo) and William Okello (National Fisheries Resources Research Institute, Jinja, Uganda), and
8
9 698 their respective teams for logistic support during the cruises for help during fieldwork. We are
10
11 699 grateful to Marc Llíros, for help with laboratory and fieldwork and the rest to the team of University
12
13 700 of Namur, Belgium. In addition we are thank to Caterina Rodriguez-Giner for their help in molecular
14
15 701 analyses and the rest of the team at the Institut de Ciències del Mar, CSIC, Barcelona. Bioinformatics
16
17 702 analyses were performed at the MARBITS platform (<http://marbits.icm.csic.es>), ICM-CSIC; as well
18
19 703 as to the two anonymous reviewers whose thorough work greatly improved the early versions of this
20
21 704 manuscript.
22
23
24 705

26 706 **Conflict of interest**

27
28 707 The authors declare no conflict of interest
29
30 708

32 709 **References**

33
34 710
35 711 Annenkova NV, Hansen G, Moestrup Ø, Rengefors K. Recent radiation in a marine and freshwater
36 712 dinoflagellate species flock. *The ISME Journal* 2015;**9**:1821-1834.
37
38 713 Annenkova NV, Rodriguez-Giner C, Logares R. Tracing the Origin of Planktonic Protists in an
39 714 Ancient Lake. *Microorganisms* 2020;**8**:543.
40
41 715 Arroyo AS, López-Escardó D, Kim E, Ruiz-Trillo I, Najle SR. Novel Diversity of Deeply Branching,
42 716 Holomycota and Unicellular Holozoans Revealed by Metabarcoding in Middle Paraná River,
43 717 Argentina. *Front Ecol Evol* 2018;**6**:99.
44
45 718 Boenigk J, Arndt H.. Bacterivory by heterotrophic flagellates: community structure and feeding
46 719 strategies. *Antonie Van Leeuwenhoek* 2002;**81**:465-80.
47
48 720 Boenigk J, Wodniok S, Bock C, Beisser D, Hempel C, Grossmann L *et al.* Geographic distance and
49 721 mountain ranges structure freshwater protist communities on a European scale. *Metab*
50 722 *Metagen* 2018;**2**:e21519.
51
52 723 Boenigk J, Wodniok S, Glücksman, E. *Biodiversity and Earth History*. Springer, 2015.
53
54 724 Bootsma HA, Hecky RE. Conservation of the African Great Lakes: A Limnological Perspective.
55 725 *Conserv Biol* 1993;**7**(3):644-656.
56
57 726 Borges AV, Abril G, Delille B, Descy JP, Darchambeau F. Diffusive methane emissions to the
58 727 atmosphere from Lake Kivu (Eastern Africa). *J Geophys Res* 2011;**116**:G03032.
59
60

- Bråte J, Logares R, Berney C, Ree DK, Klaveness D, Jakobsen KS *et al.* Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *ISME J.* 2010;**4**:1144-1153.
- Brooks JL. Speciation in ancient lakes. *Q Rev Biol* 1950;**25**(2):131-176.
- Burki F, Roger AJ, Brown MW, Simpson AG. The new tree of eukaryotes. *Trends Ecol Evol* 2020;**35**(1):3-55.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;**25**:1972-1973.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;**7**:335.
- Carlson RE. A trophic state index for lakes. *Limnol Oceanogr* 1977;**22**:361-369.
- Casamayor EO, Massana R, Benlloch S, Ovreas L, Diez B, Goddard VJ *et al.* Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. *Environ Microbiol* 2002;**4**:338-348.
- Catalan J, Camarero L, Felip M, Pla S, Ventura M, Buchaca T *et al.* High mountain lakes: extreme habitats and witnesses of environmental change. *Limnetica* 2006;**25**:551-584.
- Charvet S, Vincent WF, Comeau AM, Lovejoy C. Pyrosequencing analysis of the protist communities in a High Arctic meromictic lake: DNA preservation and change. *Front Microbiol* 2012;**3**:422.
- Chen M, Chen F, Yu Y *et al.* Genetic diversity of eukaryotic microorganisms in Lake Taihu, a large shallow subtropical lake in China. *Microb Ecol* 2008;**56**:572-83.
- Cohen AS, Kaufman L, Ogutu-Ohwayo R. Anthropogenic threats, impacts and conservation strategies in the African Great Lakes: A review. In: T. C. Johnson and E. Odada (eds.). *The Limnology, Climatology and Paleoclimatology of the East African Lakes*. Toronto: Gordon and Breach, 1996, 575-624.
- Corrêa LL, Oliveira MSB, Tavares-Dias M, Ceccarelli PS, Corrêa LL *et al.* Infections of *Hypostomus* spp. by *Trypanosoma* spp. and leeches: a study of hematology and record of these hirudineans as potential vectors of these hemoflagellates. *Rev Bras Parasitol Vet* 2016;**25**:299-305.
- Cruaud P, Vigneron A, Fradette M-S, Dorea CC, Culley AI, Rodriguez MJ, Charette SJ. Annual Protist Community Dynamics in a Freshwater Ecosystem Undergoing Contrasted Climatic Conditions: The Saint-Charles River (Canada). *Front Microbiol* 2019;**10**:2359.
- de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C *et al.* Eukaryotic plankton diversity in the sunlit ocean. *Science* 2015;**348**:1261605.
- Debroas D, Domaizon I, Humbert J-F, Jardillier L, Lepère C, Oudart A *et al.* Overview of freshwater microbial eukaryotes diversity: a first analysis of publicly available metabarcoding data. *FEMS Microbiol Ecol* 2017;**93**(4),fix023.
- Debroas D, Hugoni M, Domaizon I. Evidence for an active rare biosphere within freshwater protists community. *Mol Ecol* 2015; **24**:1236-1247.
- del Campo J, Guillou L, Hehenbergera E, Logares R, López-García P, Massana R. Ecological and evolutionary significance of novel protist lineages. *Eur J Protistol* 2016;**55**(Pt A):4-11.

- del Campo J, Massana R. Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. *Protist* 2011;**162**:435-448.
- del Campo J, Not F, Forn I, Sieracki ME, Massana R. Taming the smallest predators of the oceans. *ISME J.* 2013;**7**:351-358.
- David GM, Moreira D, Reboul G, Annenkova NV, Galindo LJ, Bertolino P *et al.* Environmental drivers of plankton protist communities along latitudinal and vertical gradients in the oldest and deepest freshwater lake. *Environ Microb* 2021;**23**(3):1436-1451.
- Descy JP, Darchambeau F, Schmid M. *Lake Kivu: Limnology and biogeochemistry of a tropical great lake*. Springer, 2012.
- Descy JP, Sarmiento H. Microorganisms of the East African Great Lakes and their response to environmental changes. *Freshwat Rev* 2008;**1**:59-73.
- Domingues CD, Silva LHS, Rangel LM, Magalhães L, de Melo RA, Lobaõ LM, Paiva R, Roland F., Sarmiento H. Microbial food-web drivers in tropical reservoirs. *Microb Ecol* 2017;**73**:505-520.
- Du P, Jiang ZB, Wang YM *et al.* Spatial heterogeneity of the planktonic protistan community in a semi-closed eutrophic bay, China. *J Plankton Res* 2019;**00**(00):1-17.
- Dyal PL, Hope S, Roberts DM, Embley TM. Use of the PCR and fluorescent-probes to recover SSU ribosomal-RNA gene-sequences from single cells of the ciliate protozoan *Spathidium*. *Mol Ecol* 1995;**4**:499-503.
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011;**27**:2194-2200.
- Fenchel T, Finlay BJ. Free-living protozoa with endosymbiotic methanogens. In Hackstein JHP (ed.). *(Endo)symbiotic Methanogenic Archaea*, Berlin: Springer Heidelberg, 2010,1-11.
- Fermani P, Torremorell A, Lagomarsino L *et al.* Microbial abundance patterns along a transparency gradient suggest a weak coupling between heterotrophic bacteria and flagellates in eutrophic shallow Pampean lakes. *Hydrobiologia* 2015;**752**:103-23.
- Filker S, Forster D, Weinisch L, Mora-Ruiz M, González B, Fariás ME, Rosselló-Mora R, Stoeck T. Transition boundaries for protistan species turnover in hypersaline waters of different biogeographic regions. *Environ Microbiol* 2017;**19**(8):3186-3200.
- Fryer G., Iles TD. *The cichlid fishes of the great lakes of Africa*. Edinburgh: Oliver and Boyd, 1972.
- Galluzzi L, Penna A, Bertozzini E, Vila M, Garcés E, Magnani M. Development of a real-time PCR assay for rapid detection and quantification of *Alexandrium minutum* (a dinoflagellate). *App. Environ Microbiol* 2004;**70**:1199-1206.
- Genitsaris S, Stefanidou N, Moustaka-Gouni M, Sommer U, Tsipas G. Variability and Community Composition of Marine Unicellular Eukaryote Assemblages in a Eutrophic Mediterranean Urban Coastal Area with Marked Plankton Blooms and Red Tides. *Diversity* 2020;**12**(3):114.
- Gile GH, Stern RF, James ER, Keeling PJ. DNA barcoding of chlorarachniophytes using nucleomorph ITS sequences. *J Phycol* 2010;**46**(4):743-750.
- Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front Microbiol* (2017); **8**:2224.

- Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner 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* (2013);**41**:D597–D604.
- Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G *et al.* Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* (2011);**21**: 494-504.
- Hammer Ø. PAST. *Paleontological Statistics. Reference Manual*. Natural History Museum, University of Oslo, 2016.
- Hecky RE, Bugenyi FWB. Hydrology and chemistry of the Great Lakes and water quality issues: problems and solutions. *Mitt. internat. Verein. Limnol.* 1992;**23**:45-54.
- Hecky RE, Fee EJ, Kling HJ, Rudd JW. Relationship between primary production and fish production in Lake Tanganyika. *Trans Am Fisher Soc* 1981;**110**:336-345.
- Hurlbert SH. The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 1971;**52**:577-586.
- İnceoğlu O, Llíros M, Crowe SA, García-Armisen T, Morana C, Darchambeau F, Borges AV, Descy JP, Servais P. Vertical distribution of functional potential and active microbial communities in meromictic Lake Kivu. *Microb Ecol* 2015;**70**:596-611.
- Ishida KI, Yabuki A, Ota S. The chlorarachniophytes: evolution and classification. *Syst Ass Vol Ser* 2007;**75**:171.
- Jobard M, Wawrzyniak I, Bronner G, Marie D, Vellet A, Sime-Ngando T, Debroas D, Lepère C. Freshwater Perkinsea: diversity, ecology and genomic information, *J Plankton Res* 2020;**42**(1):3-17.
- Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;**14**(6):587-589.
- Kammerlander B, Breiner HW, Filker S, Sommaruga R, Sonntag B, Stoeck T. High diversity of protistan plankton communities in remote high mountain lakes in the European Alps and the Himalaya mountains. *FEMS Microbiol Ecol* 2015;**91**:fiv010.
- Katoh K, Misawa K, Kuma L, Miyata Y. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;**30**:3059-3066.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772-780.
- Khomich M, Kauserud H, Logares R, Rasconi S, Andersen T. Planktonic protistan communities in lakes along a large-scale environmental gradient. *FEMS Microbiol Ecol* 2017;**93**:fiw23.
- Kilham P., Kilham SS. Endless summer: Internal loading processes dominate nutrient cycling in tropical lakes. *Freshw Biol* 1990;**23**:379-389.
- Lê S, Josse J, Husson F. FactoMineR: An RPackage for Multivariate Analysis. *J Stat Softw* 2008;**25**:1.
- Lefèvre E, Bardot C, Noël C, Carrias JF, Viscogliosi E, Amblard C, Sime-Ngando T. Unveiling fungal zooflagellates as members of freshwater picoeukaryotes: evidence from a molecular diversity study in a deep meromictic lake. *Environ Microbiol* 2007;**9**:61-71.

- 852 Lefranc M, Thénot A, Lepère C, Debroas D. Genetic diversity of small eukaryotes in lakes differing
853 by their trophic status. *Appl Environ Microb* 2005;**71**:5935-42.
- 854 Lepère C, Domaizon I, Debroas D. Unexpected importance of potential parasites in the composition
855 of the freshwater small-eukaryote community. *Appl Environ Microbiol* 2008;**74**:2940-2949.
- 856 Lepère C, Domaizon I, Hugoni M, Vellet A, Debroas D. Diversity and dynamics of active
857 microbial eukaryotes in the anoxic zone of a freshwater meromictic lake (Pavin, France).
858 *Front Microbiol* 2016;**7**:130.
- 859 Lepère C, Domaizon I, Humbert JF, Jardillier L, Hugoni M and Debroas D. Diversity, spatial
860 distribution and activity of fungi in freshwater ecosystems. *Peer J* 2019;**7**:e6247.
- 861 Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic
862 Acids Res* 2019;**47**(W1):W256-W259.
- 863 Liu J, Li X, Chai B. Effects of Seasonal Freezing-thawing on the Protistan Communities in a
864 Mountain Lake. 2021. DOI: <https://doi.org/10.21203/rs.3.rs-360259/v1>
- 865 Llíros M, Descy, JP, Libert X, Morana C, Schmitz M, Wimba L. *et al.* (2012). Microbial ecology of
866 Lake Kivu. In: J.-P. Descy, F. Darchambeau & M. Schmid (eds). *Lake Kivu: Limnology and
867 Biogeochemistry of a tropical great lake. Aquatic Ecology Series*. The Netherlands: Springer,
868 85-105.
- 869 Llíros M, Gich F, Plasencia A, Auguet JC, Darchambeau F, Casamayor EO, Descy JP, Borrego CM
870 Vertical distribution of ammonia-oxidizing crenarchaeota and methanogens in the epipelagic
871 waters of Lake Kivu. *Appl Environ Microbiol* 2010;**76**:6853-6863.
- 872 Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R. *et al.* Patterns of Rare and Abundant
873 Marine Microbial Eukaryotes. *Curr Biol* 2014;**24**:813-821.
- 874 Logares R, Audic S, Santini S, Pernice MC, de Vargas C, Massana R. Diversity patterns and activity o
875 uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *ISME J*
876 2012;**6**:1823-1833.
- 877 Loisel S, Cózar A, Adgo E, Ballatore T, Chavula G, Descy JP, Harper DM, *et al.* Decadal trends
878 and common dynamics of the bio-optical and thermal characteristics of the African Great
879 Lakes. *PLoS ONE* 2014;**9**(4):e93656.
- 880 Lopez-Garcia P, Rodriguez-Valera F, Pedros-Alio C, Moreira D. Unexpected diversity of small
881 eukaryotes in deep- sea Antarctic plankton. *Nature* 2001;**409**:603-607.
- 882 Mahé F, de Vargas C, Bass D. *et al.* Parasites dominate hyperdiverse soil protist communities in
883 Neotropical rainforests. *Nat Ecol Evol* 2017;**1**:0091.
- 884 Marquardt M, Vader A, Stübner EI, Reigstad M, Gabrielsen TM. Strong seasonality of marine
885 microbial eukaryotes in a High-Arctic Fjord (Isfjorden, in West Spitsbergen, Norway). *Appl
886 Environ Microbiol* 2016;**82**:1868-1880.
- 887 Massana R, Castresana J, Balagué V, Guillou L, Romari K, Groisillier A, Valentin K, Pedrós-Alió C
888 Phylogenetic and ecological analysis of novel marine **S**tramenopiles. *Appl Environ
889 Microbiol* 2004; **70**:3528-3534.
- 890 Massana R, Gobet A, Audic S, Bass D, Bittner L, Boutte C, Chambouvet A, Christen R. *et al.* Marine
891 protist diversity in European coastal waters and sediments as revealed by high-throughput
892 sequencing. *Environ Microbiol* 2015;**17**:4035-4049.

- 893 Massana R, Pernice M, Bunge JA, del Campo J. Sequence diversity and novelty of natural
894 assemblages of picoeukaryotes from the Indian Ocean. *ISME J* 2011;**5**:184-195.
- 895 Massana R, Terrado R, Forn I, Lovejoy C, Pedrós-Alió C. Distribution and abundance of uncultured
896 heterotrophic flagellates in the world oceans. *Environ Microbiol* 2006;**8**:1515-1522.
- 897 Metz S, Lopes dos Santos A, Castro-Berman M, Bigeard E, Licursi M, Not F, Lara E, Unrein F.
898 Diversity of photosynthetic picoeukaryotes in eutrophic shallow lakes as assessed by
899 combining flow cytometry cell-sorting and high throughput sequencing. *FEMS Microbiol*
900 *Ecol* 2019;**95**(5):fiz038.
- 901 Morana C, Roland FA, Crowe SA, Llíros M, Borges AV, Darchambeau F, Bouillon S.
902 Chemoautotrophy and anoxygenic photosynthesis within the water column of a large
903 meromictic tropical lake (Lake Kivu, East Africa). *Limnol Oceanogr* 2016;**61**:1424-1437.
- 904 Morana C, Sarmiento H, Descy JP, Gasol JM, Borges A, Bouillon S. Production of dissolved organic
905 matter by phytoplankton and its uptake by heterotrophic prokaryotes in large tropical lakes.
906 *Limnol Oceanogr* 2014;**59**:1364-1375.
- 907 Mukherjee I, Hodoki Y, Okazaki Y, Fujinaga S, Ohbayashi K, Nakano S-i Widespread. Dominance of
908 Kinetoplastids and Unexpected Presence of Diplonemids in Deep Freshwater Lakes. *Front*
909 *Microbiol* 2019;**10**:2375.
- 910 Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic
911 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;**32**(1):268-
912 274.
- 913 Obiol A, Giner CR, Sánchez P, Duarte CM, Acinas SG, Massana R. A metagenomic assessment of
914 microbial eukaryotic diversity in the global ocean. *Mol Ecol Resour* 2020;**20**(3):718-731.
- 915 Odada EO, Olago DO, Bugenyi F, Kulindwa K, Karimumuryango J, West K, Ntiba M, Wandiga S,
916 Aloo-Obudho P, Achola P. Environmental assessment of the East African Rift Valley lakes.
917 *Aquat Sci* 2003;**65**:254-271.
- 918 Ogutu-Ohwayo R, Hecky RE, Cohen AS, Kaufman L. Human impacts on the African Great Lakes.
919 *Environ Biol Fishes* 1997;**50**:117-131.
- 920 Oikonomou A, Pachiadaki M, Stoeck T. Protistan grazing in a meromictic freshwater lake with anoxic
921 bottom water. *FEMS Microb Ecol* 2014;**87**:691-703.
- 922 Okamoto N, Chantangsi C, Horák A, Leander BS, Keeling PJ. Molecular phylogeny and description
923 of the novel katablepharid *Roombia truncata* gen. et sp. nov., and establishment of the
924 Hacrobia taxon nov. *PLoS One* 2009;**4**:e7080.
- 925 Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara RB *et al*.
926 *Vegan: Community Ecology Package, R Package Version 2.5-4*, 2019.
- 927 Olapade OJ, Omitoyin BO. Anthropogenic pollution impact on physico-chemical characteristics of
928 lake Kivu, Rwanda. *African Journal of Food, Agriculture, Nutrition and Development. Afr J*
929 *Food Agric Nutr Dev*. 2012;**12**(5): 6517-6536.
- 930 Ortiz-Álvarez R, Triadó-Margarit X, Camarero L, Casamayor EO, Catalan J. High planktonic
931 diversity in mountain lakes contains similar contributions of autotrophic, heterotrophic and
932 parasitic eukaryotic life forms. *Sci Rep* 2018;**8**:4457.
- 933 Otu MK, Ramlal P, Wilkinson P, Hall RI, Hecky RE. Paleolimnological evidence of the effects of
934 recent cultural eutrophication during the last 200 years in Lake Malawi, East Africa. *J Great*
935 *Lakes Res* 2011;**37**:61-74.

- 936 Palma-Silva C, Albertoni EF, Esteves FA. Charophytes as nutrient and energy reservoir in a tropical
937 coastal lagoon impacted by humand (RJ, Brazil). *Braz J Biol Sci* 2004;**64**(3a):479-487.
- 938 Pasche N, Dinkel C, Müller B, Schmid M, Wüest A. Wehrli B. Physical and bio-geochemical limits to
939 internal nutrient loading of meromictic Lake Kivu. *Limnol Oceanogr* 2009;**54**:1863-1873.
- 940 Pernice MC, Giner CR, Logares R, Perera-Bel J, Acinas SG, Duarte CM *et al.* Large variability of
941 bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J.* 2016;**10**:
942 945-958.
- 943 Pirlot S, Vanderheyden J, Descy JP., Servais P. Abundance and biomass of heterotrophic micro-
944 organisms in Lake Tanganyika. *Freshw Biol* 2005;**50**(7):1219-1232.
- 945 Pomeroy LR. The ocean's food web, a changing paradigm. *BioScience* 1974;**24**:499-504.
- 946 Porter KG, Paerl H, Hodson R, Pace M, Priscu J, Riemann B, Scavia D., Stockner J. Microbial
947 interactions in lake food webs. In: Carpenter, S.R. (ed.). *Complex Interactions in Lake*
948 *Communities*. New York: Springer-Verlag, 1988.
- 949 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. The SILVA
950 ribosomal RNA gene database project: improved data processing and web-based tools.
951 *Nucleic Acids Res* 2013; **41**:D590-D596.
- 952 Reboul G, Moreira D, Bertolino P, Hillebrand-Voiculescu AM, López-García P. Microbial
953 eukaryotes in the suboxic chemosynthetic ecosystem of Movile Cave, Romania. *Environ*
954 *Microbiol Rep* 2018;**11**(3):464-473.
- 955 Roland FAE, Darchambeau F, Borges AV, Morana C, De Brabandere L, Thamdrup B, Crowe SA
956 Denitrification, anaerobic ammoniumoxidation, and dissimilatory nitrate reduction to
957 ammonium in an east African Great Lake (Lake Kivu). *Limnol Oceanogr* 2017;**63**:687-701.
- 958 Salani FS, Arndt H, Hausmann K, Nitsche F and Scheckenbach F. Analysis of the community
959 structure of abyssal kinetoplastids revealed similar communities at larger spatial scales. *ISME*
960 *J.* 2012;**6**:713-723.
- 961 Sandgren CD. The ecology of chrysophyte Xagellates: their growth and perennation strategies as
962 freshwater phytoplankton. In: Sandgren CD (eds). *Growth and reproductive strategies of*
963 *freshwater phytoplankton*. Cambridge: University Press Cambridge, 1988, 9-104.
- 964 Sarmiento H, Darchambeau F, Descy JP. Phytoplankton of Lake Kivu. In: Descy JP, Darchambeau F,
965 Schmid M (eds). *Lake Kivu: Limnology and biogeochemistry of a tropical great lake*.
966 Dordrecht: Springer, 2012, 67-83.
- 967 Sarmiento H, Isumbisho M, Descy JP. Phytoplankton ecology of Lake Kivu (Eastern Africa). *J*
968 *Plankton Res* 2006;**28**:815-829.
- 969 Sarmiento H, Unrein F, Isumbisho M, Stenuite S, Gasol JM, Descy JP. Abundance and distribution of
970 picoplankton in tropical, oligotrophic Lake Kivu, eastern Africa. *Freshw Biol* 2008;**53**:756-
971 771.
- 972 Sarmiento H. New paradigms in tropical limnology: the importance of the microbial food web.
973 *Hydrobiologia* 2012;**686**(1):1-14.
- 974 Schloss PD, Girard R, Martin T, Edwards, Thrash JC. The status of the microbial census: an update.
975 *MBio* 2016;**7**(3):e00201-16.
- 976 Schloss PD. Evaluating different approaches that test whether microbial communities have the same
977 structure. *ISME J* 2008;**2**:265-275.

- 978 Schmid M, Halbwachs M, Wehrli B, Wüest A. Weak mixing in Lake Kivu: new insights indicate
979 increasing risk of uncontrolled gas eruption. *Geochem Geophys Geosys* 2005;**6**:Q07009.
- 980 Schmid M, Tietze K, Halbwachs M, Lorke A, McGinnis D & Wüest A. How hazardous is the gas
981 accumulation in Lake Kivu? Arguments for a risk assessment in light of the Nyiragongo
982 volcano eruption of 2002. *Acta Vulcanol* 2002;**14**(1–2):115-122.
- 983 Schmid M, Wüest A. Stratification, mixing and transport processes in Lake Kivu. In: Descy JP,
984 Darchambeau F, Schmid M (eds). *Lake Kivu: Limnology and biogeochemistry of a tropical
985 great lake*. Dordrecht: Springer, 2012, 67–83.
- 986 Schubert H, Blindow I. *Charophytes of the Baltic Sea*. Königstein: Koeltz Scientific, 2004.
- 987 Seehausen O, Witte F, Katunzi EE, Smits J, Bouton N. Patterns of the remnant cichlid fauna in
988 southern Lake Victoria. *Conserv Biol* 1997;**11**(4):890-904.
- 989 Seehausen O. African cichlid fish: a model system in adaptive radiation research. *Proc R Soc B* 2006;
990 **273**:1987-1998.
- 991 Sherr EB, Sherr BF. Role of microbes in pelagic food webs: a revised concept. *Limnol Oceanogr*
992 1988;**33**:1225-7.
- 993 Simon M, Jardillier L, Deschamps P, Moreira D, Restoux G, Bertolino P, López-García P. Complex
994 communities of small protists and unexpected occurrence of typical marine lineages in
995 shallow freshwater systems. *Environ Microbiol* 2014;**17**:3610-27.
- 996 Simon M, López-García P, Deschamps P, Moreira D, Restoux G, Bertolino P, Jardillier L. Marked
997 seasonality and high spatial variability of protist communities in shallow freshwater
998 systems. *ISME J* 2015;**9**(9):1941-1953.
- 999 Simon M, López-García P, Deschamps P, Restoux G, Bertolino P, Moreira D *et al*. Resilience of
1000 freshwater communities of small microbial eukaryotes undergoing severe drought events.
1001 *Front Microbiol* 2016;**7**:812.
- 1002 Singer D, Seppely CVW, Lentendu G *et al*. Protist taxonomic and functional diversity in soil,
1003 freshwater and marine ecosystems. *Environ Int* 2021;**146**:106262.
- 1004 Spigel RH, Coulter GW. Comparison of hydrology and physical limnology of the East African Great
1005 Lakes: Tanganyika, Malawi, Victoria, Kivu and Turkana (with reference to some North
1006 American Great Lakes). In: T.C. Johnson and E. Odada (eds.) *The Limnology, Climatology
1007 and Paleoclimatology of East African Lakes*. Toronto, Gordon and Breach, 1996, 103-139.
- 1008 Stenuite S, Tarbe AL, Sarmiento H, Unrein F *et al*. Photosynthetic picoplankton in Lake Tanganyika:
1009 biomass distribution patterns with depth, season and basin. *J Plankton Res* 2009;**31**:1531-
1010 1544.
- 1011 Stoeck T, Bass D, Nebel M, Christen R, Jones MD, Breiner HW *et al*. Multiple marker parallel tag
1012 environmental DNA sequencing reveals a highly complex eukaryotic community in marine
1013 anoxic water. *Mol Ecol* 2010;**19**(1): 21-31.
- 1014 Tarbe AL, Stenuite S, Balagué V, Sinyinza D, Descy JP, Massana R. Molecular characterization of
1015 the small-eukaryote community in a tropical Great Lake (Lake Tanganyika, East Africa).
1016 *Aquat Microb Ecol* 2011b;**62**:177-190.
- 1017 Tarbe AL, Unrein F, Stenuite S, Pirlot S, Sarmiento H, Sinyinza D, Descy JP. Protist herbivory: a key
1018 pathway in the pelagic food web of Lake Tanganyika. *Microb Ecol* 2011a;**62**:314-323.

- 1019 Tarbe AL. Les protistes: acteurs-clés du réseau trophique pélagique au Lac Tanganyika. PhD thesis,
1020 University of Namur. 2010.
- 1021 Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D Eukaryotic microbes, species
1022 recognition and the geographic limits of species: examples from the kingdom Fungi. *Philos*
1023 *Trans R Soc Lond B Biol Sci* 2006;**361**:1947-1963.
- 1024 ter Braak CJF, Smilauer P. CANOCO reference manual and user's guide to Canoco for Windows—
1025 software for canonical community ordination (version 4). Ithaca: Microcomputer Power,
1026 1998.
- 1027 ter Braak CJF. Ordination. In Jongman RHG, ter Braak CJF, van Tongeren OFR (eds). *Data analysis*
1028 *in community and landscape ecology*. Wageningen, Cambridge University Press, 1987, 91-
1029 173.
- 1030 Triadó-Margarit X., Casamayor EO. Genetic diversity of planktonic eukaryotes in high mountain
1031 lakes (Central Pyrenees, Spain). *Environ Microbiol* 2012;**14**:2445-2456.
- 1032 Vadeboncoeur Y, McIntyre PB, Vander Zander MJ. Borders of Biodiversity: Life at the Edge of the
1033 World's Large Lakes. *BioScience* 2011;**61**:7.
- 1034 Verleyen E, Vyverman W, Sterken M, Hodgson DA, De Wever A, Juggins S, Van de Vijver B, Jones
1035 VJ, Vanormelingen P, Roberts D. The importance of dispersal related and local factors in
1036 shaping the taxonomic structure of diatom metacommunities. *Oikos* 2009;**118**:1239-1249.
- 1037 Weiss RF. Determinations of carbon dioxide and methane by dual catalyst flame ionization
1038 chromatography and nitrous oxide by electron capture chromatography. *J Chromatogr Sci*
1039 1981;**19**:611-616.
- 1040 Weisse T, Müller H, Pinto-Coelho RM, Schweizer A, Springmann D, Baldringer G. Response of the
1041 microbial loop to the phytoplankton spring bloom in a large prealpine lake. *Limnol Oceanogr*
1042 1990;**35**:781-794.
- 1043 Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. Rethinking the
1044 marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science*
1045 2015;**347**:1257594.
- 1046 Xiong W, Jousset A, Li R *et al.* A global overview of the trophic structure within microbiomes across
1047 ecosystems. *Environ Int* 2021;**151**:106438.
- 1048 Xu JP, Li F, Zhang B *et al.* Estimation of chlorophyll-a concentration using field spectral data: a case
1049 study in inland Case-II waters, North China. *Environ Monit Assess* 2009;**158**:105-116.
- 1050 Yubuki N, Leander BS. Diversity and Evolutionary History of the Symbiontida (Euglenozoa). *Front*
1051 *Ecol Evol* 2018;**6**:100.
- 1052 Zhao F, Filker S, Xu K, Huang P, Zheng S. Patterns and drivers of vertical distribution of the ciliate
1053 community from the surface to the abyssopelagic zone in the western pacific ocean. *Front*
1054 *Microbiol* 2017;**8**:2559.

208x191mm (300 x 300 DPI)

Fig. 2

204x285mm (300 x 300 DPI)

Fig. 3

165x278mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

Figure legends

Figure 1. (A) Map of East African Great Lakes sampled during rainy season. The left panel shows the four lakes amplified: Albert, Edward, Victoria and the two basins of Kivu: Ishungu and Gisenyi. Black circles indicate sampling sites. **(B)** Percentage total of protists high-rank supergroups found in East Great African Lakes. Total OTUs (n=779), Total OTUs Novelty (n=316). Star symbols denote taxa that were considered as supergroups in the recent classification based on new phylogenetic analyses, and circle shows taxa that is still without a clear resolution (Burki *et al.* 2020).

Figure 2. Vertical profiles of the main physical-chemical parameters of Lake Kivu at the Ishungu basin (A,C) and the Gisenyi basin (B,D). (A,B) Temperature ($^{\circ}\text{C}$), DO: dissolved oxygen (mg L^{-1}) and pH. (C,D) Cond: conductivity ($\mu\text{S cm}^{-1}$), CH_4 : methane (μM) and NO_x : oxidized nitrogen (μM).

Figure 3. Novelty pattern derived from OTUs among supergroups. Dots represent the percentage of similarity with the closest environmental match (CEM) and the closest cultured match (CCM) for each OTU within each supergroup. Bars represents abundance frequencies.

Figure 4. Community structure and taxonomic composition of surface protistan assemblages in the four East Great African Lakes (two sites in Kivu). (A) Relative proportion of reads within taxonomic groups. For a better interpretation, the graph was made with groups that presented > 5 reads in all the samples. (B) Principal component analyses (PCA) based on microbial communities interpreting the variation with the supplementary physico-chemical parameters. Samples appear in solid red circle and each arrow points in the direction of the steepest increase of the corresponding values. Taxonomic groups enriched in a given part of the plot are shown. Albert -A-, Victoria -V- and Edward -E-: 0.8-10 μm ; I-5 and G-5 -Ishungu and Gisenyi at 5 m-: 0.2-5 μm . (Temp: water temperature, Cond: conductivity, Chl-*a*: chlorophyll *a*, DIN: dissolved inorganic nitrogen, SRP: soluble reactive phosphorous, DO: dissolved oxygen, Depth: maximum depth). (C) Venn diagram showing OTUs shared by the five surface samples.

Figure 5. Community structure and taxonomic composition of protistan assemblages over vertical profiles of Lake Kivu. **(A)** Ishungu, **(B)** Gisenyi. For a better interpretation histograms were made with taxonomic groups that presented > 5 reads in all the samples. **(C)** Principal component analyses (PCA) based on microbial communities interpreting the variation with the supplementary physico-chemical parameters. Samples appear in solid red circles and each arrow points to the direction of the steepest increase of the corresponding values. Taxonomic groups enriched in a given part of the plot are shown. Ishungu: I-5, I-15, I-30, I-45, I-50 and I-70. Gisenyi: G-5, G-15, G-25, G-35 and G-70. (Temp: water temperature, Cond: conductivity, Chl-*a*: chlorophyll *a*, DO: dissolved oxygen, NO₃⁻: nitrates).

Figure 6. Phylogenetic tree and classification of Stramenopiles OTUs. From the phylogenetic tree, circles in red correspond to bootstrap support values > 75%. Clusters highlighted in orange correspond to environmental clusters that were formed by novel OTUs. The collapsed clusters were annotated with the number of OTUs followed by the number of novel OTUs and the number of reference sequences (rs). The right panel shows the number of reads from each OTU in each sample from Ishungu (I-5, I-15, I-30, I-45, I-50 and I-70) and Gisenyi (G-5, G-15, G-25, G-35 and G-70). The circles represent the presence of the OTU in the corresponding sample and the values inside are the number of reads.

Figure 7. Phylogenetic tree and classification of Alveolata OTUs. From the phylogenetic tree, circles in red correspond to bootstrap support values > 75%. Clusters highlighted in yellow correspond to environmental clusters that were formed by novel OTUs. The collapsed clusters were annotated with the number of OTUs followed by the number of novel OTUs and the number of reference sequences (rs). The right panel shows the number of reads from each OTU in each sample from Ishungu (I-5, I-15, I-30, I-45, I-50 and I-70) and Gisenyi (G-5, G-15, G-25, G-35 and G-70). The circles represent the presence of the OTU in the corresponding sample and the values inside are the number of reads.

Figure 8. Phylogenetic tree and classification of ciliates OTUs. From the phylogenetic tree, circles in red correspond to bootstrap support values > 75%. Clusters highlighted in yellow correspond to environmental clusters that were formed by novel OTUs. The collapsed clusters were annotated with the number of OTUs followed by the number of novel OTUs and the number of reference sequences (rs). The right panel shows the number of reads from each OTU in each sample from Ishungu (I-5, I-15, I-30, I-45, I-50 and I-70) and Gisenyi (G-5, G-15, G-25, G-35 and G-70). The circles represent the presence of the OTU in the corresponding sample and the values inside are the number of reads.

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. Main geographical and limnological characteristics of East African Great Lakes.

Lake	Kivu Ishungu	Kivu Gisenyi	Edward	Albert	Victoria
	(I)	(G)	(E)	(A)	(V)
Coordinates	02°16'S, 28°59'E	01°47'S, 29°12'E	00°12'N, 29°49'E	01°98'N, 31°16'E	00°39'N, 33°16'E
Altitude (msnm)	1463	1481	928	620	1151
Lake Area (Km ²)	2322 (main lake)	2322 (main lake)	2325*	5300*	68800*
Maximum Depth (m)	160 [†]	440 [†]	117*	58*	79*
Mean euphotic depth 1% (m)	18 (main lake)*	18 (main lake)*	13*	12*	9*
Mixing Regime	Meromictic	Meromictic	Holomictic	Holomictic	Holomictic

*Morana *et al.* 2014, [†]Pasche *et al.* 2009

Table 2. Main physical, chemical and biological variables of East African Great Lakes in surface mixed-layer waters.

Lake	Kivu Ishungu (I)	Kivu Gisenyi (G)	Edward (E)	Albert (A)	Victoria (V)
Temperature (°C)	24.10	24.30	26.60	27.90	25.10
pH	9.50	8.40	9.10	8.90	8.30
Conductivity ($\mu\text{S cm}^{-1}$)	1155.00	1085.50	849.20	640.70	95.00
Alkalinity (nmol Kg^{-1})	12.70	12.70	8.20	5.80	0.80
DO (mg L^{-1})	5.70	7.50	7.60	8.50	7.20
DIN (μM)	0.36	0.13	2.42	2.74	3.88
SRP (μM)	0.20	0.20	0.43	0.19	0.82
Chl- <i>a</i> (μgL^{-1})	2.10	2.10	9.90	5.90	6.00
Trophic State (TSI)	Oligotrophic	Oligotrophic	Meso- Eutrophic	Mesotrophic	Mesotrophic

DO: dissolved oxygen, DIN: dissolved inorganic nitrogen, SRP: Soluble reactive phosphorus, Chl-*a*: Chlorophyll-*a*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 3. α -diversity indices of East African Great Lakes in surface mixed-layer protistan assemblages.

Sample	Kivu Ishungu (I-5)	Kivu Gisenyi (G-5)	Edward (E)	Albert (A)	Victoria (V)
Richness (OTUs numbers_S)	122	96	113	98	125
Diversity (Simpson 1-D)	0.91	0.85	0.88	0.92	0.93
Diversity (Shannon_H)	3.35	2.97	2.98	3.30	3.59
Evenness_e^H/S	0.23	0.20	0.17	0.27	0.29

Kivu Ishungu: 5m (I-5) (0.2-5 μm), Kivu Gisenyi: 5m (G-5) (0.2-5 μm), Edward (E) (0.8-10 μm), Albert (A) (0.8-10 μm), Victoria (V) (0.8-10 μm).

Table 4: α -diversity indices of Lake Kivu vertical protistan assemblages

Sample	I-5	I-15	I-30	I-45	I-50	I-70	G-5	G-15	G-25	G-35	G-70
Richness (OTUs numbers_S)	133	146	117	146	149	216	104	150	120	110	207
Diversity (Simpson 1-D)	0.90	0.93	0.89	0.86	0.95	0.98	0.85	0.92	0.92	0.83	0.97
Diversity (Shannon_H)	3.31	3.54	2.92	3.30	3.67	4.45	2.94	3.84	3.23	2.73	4.13
Evenness_e^H/S	0.21	0.24	0.16	0.18	0.26	0.40	0.18	0.22	0.21	0.14	0.30

Ishungu: I-5, I-15, I-30, I-45, I-50 and I-70. Gisenyi: G-5, G-15, G-25, G-35 and G-70.

Supplementary Material

Microbial eukaryotes assemblages and potential novel diversity in four tropical East-African Great Lakes

Paulina Fermani*, Sebastián Metz, Vanessa Balagué, Jean Pierre Descy, Cédric Morana, Ramiro Logares, Ramon Massana, Hugo Sarmento

*Corresponding author: pfermani@intech.gov.ar

Supplementary Table S1. Primers utilized in the 15 samples of East-African Great Lakes.

Sample	Primer F	Composition	Adaptor A	Barcode	Sequence 5' - 3'	454_Primer (5'-3')
G-5	1	5'-Akey-8R1-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AACAACAA	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACAAACAAACAGCASCYGGGTAATTCC
G-15	5	5'-Akey-8R5-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCATGCG	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCATGCGCCAGCASCYGGGTAATTCC
G-25	8	5'-Akey-8R8-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTTCTTCA	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTCTTCAACAGCASCYGGGTAATTCC
G-30	10	5'-Akey-8R10-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AACAATGG	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACAAATGGCCAGCASCYGGGTAATTCC
G-35	12	5'-Akey-8R12-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GAGTACTA	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGAGTACTAACAGCASCYGGGTAATTCC
G-70	15	5'-Akey-8R15-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TATCACAT	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTATCACATCCAGCASCYGGGTAATTCC
I-5	20	5'-Akey-8R20-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CCAGTCAG	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCAGTCAGCCAGCASCYGGGTAATTCC
I-15	23	5'-Akey-8R23-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTATAAGT	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTATAAGTCCAGCASCYGGGTAATTCC
I-30	26	5'-Akey-8R26-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ATGTATAA	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGATGTATAAACAGCASCYGGGTAATTCC
I-45	31	5'-Akey-8R31-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AACGAGTG	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACGAGTGCCAGCASCYGGGTAATTCC
I-50	35	5'-Akey-8R35-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGATCCTA	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGATCCTAACAGCASCYGGGTAATTCC
I-70	36	5'-Akey-8R36-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGCACCTT	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGCACCTTCCAGCASCYGGGTAATTCC
A	38	5'-Akey-8R38-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ATACATGG	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGATACATGGCCAGCASCYGGGTAATTCC
E	40	5'-Akey-8R40-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AACAGACA	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACAGACACAGCASCYGGGTAATTCC
V	43	5'-Akey-8R43-primer-3'	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCCATGT	CCAGCASCYGGGTAATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCCATGTCCAGCASCYGGGTAATTCC

Supplementary Table S2. OTUs affiliating to MALVs (Syndiniales) recovered after phylogenetic analysis in Lake Kivu.

OTU	Group	Clades
OTU_1247	Syndiniales Dino-Group-I	Clade-1
OTU_2569	Syndiniales Dino-Group-I	Clade-1
OTU_761	Syndiniales Dino-Group-I	Clade-5
OTU_1961	Syndiniales Dino-Group-I	Clade-5
OTU_2533	Syndiniales Dino-Group-II	
OTU_776	Syndiniales Dino-Group-II	Clade-3
OTU_455	Syndiniales Dino-Group-II	Clade-3
OTU_103	Syndiniales Dino-Group-II	Clade-6
OTU_1373	Syndiniales Dino-Group-II	Clade-7
OTU_1381	Syndiniales Dino-Group-II	Clade-8
OTU_2074	Syndiniales Dino-Group-II	Clade-10-and-11
OTU_719	Syndiniales Dino-Group-II	Clade-10-and-11
OTU_158	Syndiniales Dino-Group-II	Clade-10-and-11
OTU_2110	Syndiniales Dino-Group-II	Clade-16
OTU_1002	Syndiniales Dino-Group-III	

1
2
3 43 **Supplementary Figure S1. (A)** Total read and potential "novel" abundances in surface mixed-layer waters of East
4 African Great Lakes, sampled during rainy season: Kivu (K), Edward (E), Albert (A) and Victoria (V) and depth of
5
6 44 Ishungu: I-5, I-15, I-30, I-45, I-50, I-70 and Gisengyi: G-5, G-15, G-25, G-35, G-70. **(B)** Rarefaction analysis of
7
8 45 OTUs based on 99% identity.
9
10 46 OTUs based on 99% identity.

28
29
30 48
31
32 49
33
34 50
35
36 51
37
38 52
39
40 53
41
42 54
43 55
44
45 56
46
47 57
48
49 58
50
51 59
52
53 60
54
55 61
56
57 62
58
59 63
60 64

Supplementary Figure S2. Percentage total of groups of protists per supergroups found in East Great African Lakes.

1
2
3
4 81 **Supplementary Figure S3.** Relative abundance of the most abundant OTU (relative abundance $\geq 1\%$). (A) Surface
5 82 mixed-layer waters. (B,C) Vertical profiles of Lake Kivu: Ishungu (B) and Gisenyi (C). The remaining OTUs
6 83 (relative abundance $< 1\%$) are shown in green.