Planktonic protistan communities in lakes along a large-scale environmental gradient

Maryia Khomich[1,2*], Håvard Kauserud[2], Ramiro Logares[3], Serena Rasconi[4], Tom Andersen[1]

[1]Section for Aquatic Biology and Toxicology, Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway

[2]Section for Genetics and Evolutionary Biology, Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway

[3]Department of Marine Biology and Oceanography, Institute of Marine Sciences, CSIC, 08003 Barcelona, Spain

[4]WasserCluster–Biological Station Lunz, Inter-university Centre for Aquatic Ecosystem Research, A-3293 Lunz am See, Austria

[*]Corresponding author: Maryia Khomich, Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway

Tel.: +47-22845979, Fax: +47-22854726

E-mail address: maryia.khomich@ibv.uio.no, marykhomich@gmail.com

Running title: Molecular reassessment of freshwater biodiversity

Abstract

Despite their obvious importance, our knowledge about the eukaryotic microbial diversity of inland waters is still limited and poorly documented. We applied 18S rDNA amplicon sequencing to provide a comprehensive analysis of eukaryotic diversity in 74 low-productivity lakes along a 750 km longitudinal transect (5.40-18.52ºE) across southern Scandinavia. We detected a wide diversity of pelagic microbial eukaryotes, classified into 1882 operational taxonomic units (OTUs). The highest OTU richness was found in traditional phytoplankton groups like Dinoflagellata, Chrysophyceae, Chlorophyta and Cryptophyta. A total of 53.6% OTUs were primarily autotrophic while 19.4% of the heterotrophic OTUs belonged to putative parasitic taxa. Except for a longitudinal trend in the relative influence of mixotrophs, there were no significant associations between major functional groups (auto-, heterotrophs and parasites) and spatial and environmental variables. Community dissimilarity increased significantly with increasing geographical distance between lakes. In accordance with earlier, microscopy-based surveys in this region, we demonstrate distinct gradients in protistan diversity and community composition, which are better explained by spatial structure than local environment. The strong association between longitude and protistan diversity is probably better explained by differences in regional species pools due to differences in landscape productivity than by dispersal limitation or climatic constraints.

Keywords: planktonic, protists, oligotrophic lakes, diversity, 454 pyrosequencing, longitudinal gradient

Introduction

Freshwater environments contain only around 0.01% of the world's water and comprise <1% of the land surface, but contain a disproportionately high fraction of the global biodiversity (e.g. 1/3 of global vertebrate diversity) (Dudgeon *et al.*, 2006, Strayer & Dudgeon, 2010). However, knowledge of the total diversity of inland waters is incomplete, in particular among invertebrates and microbes (Dudgeon *et al.*, 2006). Freshwater habitats are of vital importance to global element cycling (Cole *et al.*, 2007) and other ecosystem services (Aylward *et al.*, 2005). "Lakes are particularly appealing subjects for ecological study in that they are self-contained ecosystems, discrete and largely isolated from others" (Pianka, 1974). Given their discrete nature, lakes may be good model systems for evaluating the effects of local dispersal and regional productivity on meta-community properties (Leibold *et al.*, 2004, Ptacnik *et al.*, 2010, Hortal *et al.*, 2014). There is increasing awareness of how dispersal in spatial meta-communities contributes to maintaining biodiversity in freshwater ecosystems (De Meester *et al.*, 2005, Heino, 2013). Dispersal mechanisms are better understood in fish and macroinvertebrates (Shurin *et al.*, 2009) than in small microbial eukaryotes (protists) and bacteria. However, recent studies support dispersal limitation rather than cosmopolitan distribution among microbes (Martiny *et al.*, 2006, Heino *et al.*, 2010, Soininen *et al.*, 2011, Stomp *et al.*, 2011, Soininen, 2012, Bates *et al.*, 2013).

Protists are key components in aquatic food webs both as producers of organic matter and as major consumers of bacterial biomass (Šlapeta *et al.*, 2005). Recent molecular studies from marine (Logares *et al.*, 2012, Massana *et al.*, 2014, de Vargas *et al.*, 2015) and freshwater (Triadó‐Margarit & Casamayor, 2012, Mangot *et al.*, 2013) ecosystems suggest these environments represent a wide range of ecological niches and harbour enormous eukaryotic diversity. Phytoplankton monitoring has often focused on harmful blooms caused by

anthropogenic nutrient pollution (Padisak *et al.*, 2006, O'Neil *et al.*, 2012) while relatively fewer studies have systematically investigated the general protistan diversity of oligotrophic lakes.

High-throughput sequencing (HTS) allows exploration of biodiversity in complex microbial communities in greater detail than hitherto possible (Johnson & Martiny, 2015). Recently, HTS has extensively been applied in various aquatic microbial diversity surveys focusing on compositional changes along trophic or salinity gradients (Balzano *et al.*, 2015), as well as seasonal fluctuations in microbial communities (Nolte *et al.*, 2010, Simon *et al.*, 2015). Freshwater eukaryotic assemblages have also been described in meromictic lakes (Charvet *et al.*, 2012), high mountain lakes (Kammerlander *et al.*, 2015), saline lakes (Wang *et al.*, 2014) and lakes of differing trophic status (Mangot *et al.*, 2013).

A set of lakes in a large region (e.g. boreal lakes) is an ideal model to study compositional variation from a perspective of multiple communities connected by dispersing organisms (Leibold *et al.*, 2004, Hortal *et al.*, 2014). Here, we investigate planktonic protistan communities over the west-east biodiversity gradient through southern Norway and Sweden (Ptacnik *et al.*, 2008b, Ptacnik *et al.*, 2010). Lakes for this study were carefully selected to be as similar as possible with respect to properties other than spatial position and local productivity (Fig. 1, Table S1). Ptacnik *et al.* (2008b, 2010) have found distinct regional trends in unicellular plankton diversity across Scandinavian lakes. Moreover, there are clear indications that plankton diversity actually affects functional aspects of lake ecosystems, such as resource use efficiency (Ptacnik *et al.*, 2008b). The Scandinavian diversity gradient is complex and not fully resolved as it coincides both with major changes in altitude, soil depth and landscape productivity, as well as the main post-glacial dispersal routes for freshwater organisms. Recurring glaciations in boreal areas can be considered an important, though neglected, historical climatic factor influencing biota

(Soininen, 2012). Re-colonization of Fennoscandia (Norway, Sweden, Finland and Karelia), which was entirely glaciated until approximately 15 000 years ago, could occur either from the east (Russia), or from the south (Kontula & Väinölä, 2001). Molecular studies show fish (*Perca fluviatilis*) and crustacean (*Gammarus lacustris*) populations have invaded Fennoscandia from both directions (Refseth *et al.*, 1998, Vainio & Väinölä, 2003).

Local environmental effects must be controlled for to resolve spatial patterns in lake biodiversity. The concentration of total phosphorus (TP) is often the primary limiting factor for primary production in lakes (Schindler, 1977), especially in areas with high atmospheric deposition of inorganic nitrogen such as Southern Scandinavia (Elser *et al.*, 2009, Hessen *et al.*, 2009). The concentration of total organic carbon (TOC) affects the balance between heterotrophic and autotrophic processes in plankton communities by being both a carbon source for heterotrophic bacteria and a modulator of the underwater light availability (Jansson *et al.*, 2000, Thrane *et al.*, 2014). Biotic factors (predation, host availability for parasites, viral infection) also influence protistan diversity (Lepère *et al.*, 2006, Rasconi *et al.*, 2011, Zhao *et al.*, 2011, Triadó‐Margarit & Casamayor, 2012).

The phytoplankton of lakes in this region is well studied by traditional microscopy (Rosén, 1981, Brettum, 1989, Brettum & Andersen, 2004, Willén, 2007, Ptacnik *et al.*, 2010), but this is the first comprehensive HTS survey of their protistan communities. Our objectives were: (i) to analyse taxonomic and functional composition of pelagic protistan communities across a known biodiversity gradient; (ii) to determine what factors govern protistan community differentiation across lakes in this gradient; and (iii) to estimate the relative roles of local and regional factors in influencing community structure.

Materials and methods

Site description

Lakes were selected from the "Rebecca" (Solheim *et al.*, 2008) and "Nordic lake survey 1995" (Henriksen *et al.*, 1998) data sets on Norwegian and Swedish lakes to create a subset fulfilling the following criteria: longitude 5 – 18 °E, latitude 58 – 62 °N, altitude < 600 m, surface area > 1 km^2, TP <30 μg L^{-1}, TOC <30 mg L^{-1} and pH > 5. These lakes represent a subset of boreal lakes with best possible coverage and orthogonality with respect to gradients of TP, TOC and longitudinal position. The two former represent two major effects on aquatic productivity (Thrane *et al.*, 2014), while the latter reflects the regional diversity gradient (Ptacnik *et al.*, 2010). This means that longitude will also be aligned with the regional productivity gradient, which Ptacnik *et al.* (2010) have shown can be represented by a distance weighted average of TP in nearby lakes from an independent data set (TPreg). Longitude and TPreg were closely correlated in our study lakes with Pearson correlations from 0.93 to 0.97, for interpolation ranges from 100 to 500 km (Fig. S1). The three gradient variables were split in two factor levels (high/low), giving eight combinations of TP, TOC and longitude. Twelve lakes were randomly sampled from each combination. Sampling was performed by hydroplane in July to August 2011 (Thrane *et al.*, 2014). Due to unfavorable weather conditions during sampling the number of sampled lakes was eventually reduced to 77 (Fig. 1).

Sampling program

Water samples were collected from the lake epilimnion (0-5 m) using an integrating water sampler (Hydro-BIOS, Germany) in the central part of each lake during daytime. For DNA analysis, up to 15 L of water was pre-screened on 100 μm mesh to remove large non-protistan

organisms and filtered onto 47 mm 2 μm Isopore TTTP membrane filters (Millipore Corp., MA, USA) taken in 3x3 replicates. The filters were stored at -20°C in cryovials until DNA extraction. Samples for nutrients were collected as described in Thrane *et al.* (2014). Concentrations of TP, TOC and total nitrogen (TN) were determined using standard techniques (for details see in Thrane *et al.*, 2014).

Chemical characteristics of the water (e.g. nutrients, pH and ionic strength) are the most relevant environmental factors determining changes in phytoplankton community composition. TOC and TP were chosen as proxies to represent regional environmental gradients and local nutrient supply variability. The third variable, conductivity, is directly related to the concentration of ions in dissolved salts, and serves as an indicator of soil depth and landscape productivity that is less affected by local pollution than TP (Ryder, 1982). Since there is a close relationship between pH and conductivity in this data set ($R^2 = 0.65$; $p < 0.00001$), conductivity can also be considered a proxy for pH. It is important to take into account that not all predictor variables are completely independent (Fig. S2).

DNA extraction, amplification and 454-sequencing of the V4 SSU

DNA was extracted from the filters using NucleoSpin® Plant II Kit (Mackerey-Nagel, Düren, Germany) according to the manufacturer's instructions and quantified using Nanodrop (NanoDrop Technologies Inc, DE, USA). The hypervariable V4 region (~380 bp) of the eukaryotic 18S rRNA (Stoeck *et al.*, 2010, Logares *et al.*, 2012) was amplified using universal primers. It is the gene's longest variable region and has relatively high taxonomic resolution (Dunthorn *et al.*, 2012). Fusion pyrosequencing primers were designed according to Roche specifications and included adaptors, (Adaptor A (5'-3'):

158 CCATCTCATCCCTGCGTGTCTCCGACTCAG, adaptor B (5'-3')
159 CCTATCCCCTGTGTGCCTTGGCAGTC), key (TCAG) and 10-bp unique tags (MIDs in
160 Roche technical bulletin 005-2009) and the V4 primers. PCR amplifications were performed on a
161 PTC-200 DNA Engine Cycler (BioRad, USA) in 20-µl reaction volumes containing 4 µl DNA
162 template (diluted 1:10), 1x Phusion HF buffer, 0.2 mM dNTPs, 0.25 µM of each fusion primer,
163 0.02 U/µl Phusion HotStart II polymerase (Finnzymes, Vantaa, Finland), 3% DMSO and 1 mg
164 ml^{-1} BSA (New England BioLabs, Auckland, New Zealand). The amplification program was as
165 follows: 30 s at 98°C, followed by 30 cycles of 10 s at 98°C, 30 s at 53°C and 30 s at 72°C, with
166 a final extension step at 72°C for 5 min before storage at -20°C. Amplification was verified on
167 1% agarose gels. PCR products were cleaned with a Wizard® SV Gel and PCR Clean-Up
168 System (Promega, Madison, Wisconsin, USA), quantified using a Sequalprep™ Normalization
169 Plate (96) Kit (Invitrogen, Paisley, United Kingdom) and pooled into equimolar amplicon
170 libraries. Pooled libraries were quantified using Qubit dsDNA BR Assay Kit (Invitrogen). The
171 454 Titanium sequencing of the tagged amplicons was performed using GS FLX Titanium
172 (Lib A chemistry kit) at the Norwegian Sequencing Centre at the University of Oslo (Norway)
173 on 1/2 of a 454 FLX Titanium sequencing plate (454 Life Sciences, Branford, CT, USA). The
174 raw 454 data with corresponding mapping files were deposited in Dryad
175 (doi:xx.xxxx/dryad.xxxxx).
176
177 *Bioinformatics analyses for 454 reads*
178 A total of 526 390 sequence reads from 87 samples were quality-filtered, denoised, and
179 processed using QIIME v. 1.5.0 (Caporaso *et al.*, 2010) on the Abel cluster at the University of
180 Oslo unless otherwise indicated. All reads with mismatched forward and/or reverse tags were

181 removed to avoid false positives (Carlsen *et al.*, 2012). Sequences with length <200 bp and >550
182 bp, average Phred quality score of <25, errors in the tags, homopolymers exceeding 6 bp,
183 ambiguous base calls >1, and >1 mismatch in the primers, were discarded. Additionally, a 50-bp
184 sliding window (average quality score >25) was used to identify regions of low sequence quality
185 and sequences were truncated to the last good window. The resulting sequences (414 679) were
186 denoised using DeNoiser v. 091 (Reeder & Knight, 2010), and clustered into operational
187 taxonomic units (OTUs) with UCLUST v.1.2.22 (Edgar, 2010) with a 99% similarity threshold, -
188 -max_accepts=20, and --max_rejects=500. A high clustering threshold was used to allow for
189 inclusion of highly related but distinct taxa (e.g. among ciliates, haptophytes) (Worden, 2006,
190 Doherty *et al.*, 2007, Egge *et al.*, 2013, Santoferrara *et al.*, 2014), and because the V4 region is
191 characterized by rapid rates of evolution, and the dataset was denoised (Logares *et al.*, 2014).
192 OTUs globally represented by a single sequence were considered sequencing errors and removed
193 (Quince *et al.*, 2009, Kunin *et al.*, 2010, Tedersoo *et al.*, 2010). Taxonomic assignments were
194 made by comparing the most abundant (representative) sequence of each OTU against reference
195 databases SILVA v111 (Quast *et al.*, 2012) and PR2 (Guillou *et al.*, 2012) using BLAST
196 (Altschul *et al.*, 1990) with threshold e-value=10^{-5}. As taxonomy was consistent across both
197 databases, the SILVA assignments were used in downstream analyses. Unwanted OTUs (e.g.
198 Metazoa, Embryophyta) were removed. Chimeras were detected by using ChimeraSlayer (Haas
199 *et al.*, 2011), as implemented in mothur v.1.26.0 (Schloss *et al.*, 2009), and subsequently
200 discarded. The final, curated dataset comprised 334 858 reads (64% of reads), including 10
201 technical replicates to check for sequencing consistency. The 10 technical replicates were
202 significantly more similar with respect to OTU composition than between sample-comparisons
203 (Fig. S3), demonstrating little influence of biases introduced during PCR and sequencing.

204

205 *Statistical analyses*

206 One glacially influenced lake (temperature <10°C) and the two with low/high pH (<6 and >8)
207 were omitted from the final set of lakes as potential outliers leaving 74 lakes with pH 6.3-8.0.
208 OTUs with <10 reads or occurring in <2 samples were removed. A total of 281 571 sequences
209 (54% of initial raw reads) that clustered into 1882 OTUs for the 74 lake samples were used in
210 statistical analyses. The OTU table was rarefied to a common sampling depth of 1000
211 reads/sample to calculate OTU-based diversity measures. To minimize the effect of abundance
212 measure inconsistencies, ordinations were conducted on presence/absence data as well as by-site
213 normalized read abundances. Downstream statistical analyses were performed in R version 3.1.0
214 (R Development Core Team, 2014) using vegan (Oksanen *et al.*, 2013) and MASS (Venables &
215 Ripley, 2002) for multivariate and species richness analyses unless otherwise noted. Species
216 accumulation curves (SAC; calculated using the analytical version of the *specaccum* function)
217 were used to assess sampling effort. Rarefaction curves were constructed using *rarecurve*. Alpha
218 diversity indices (observed richness, Shannon diversity, Simpson diversity) were calculated
219 using the function *diversity*.

220 Ordinations by detrended correspondence analysis (DCA) (Hill & Gauch Jr, 1980) and non-
221 metric multidimensional scaling (NMDS) ordinations (Minchin, 1987) were used to describe
222 patterns in eukaryotic species composition. In addition, NMDS ordinations were conducted on a
223 subset of a matrix representing 10 technical replicates to confirm that sequencing-induced
224 variation was smaller than biological variation in the samples. Kendall's rank correlation
225 coefficients τ and procrustes correlations (*protest* function in vegan) between pairs of DCA and
226 NMDS axes with two and three dimensions were calculated. Permutation-based significance

tests by the *envfit* function were used to fit spatial and environmental gradient variables to the NMDS ordination. Since OTU richness is a count variable, we used generalized linear models (GLMs) of the quasi-poisson family with permutation-based significance tests to fit rarefied richness (both total and for individual taxonomic groups) to the NMDS ordination axes. A standard Mantel test to investigate correlation between community composition and geographical distance between lakes was run using Bray-Curtis distances between communities and 999 permutations.

Linear models (LMs) related OTU richness (observed, rarefied) and diversity (Shannon, Simpson) to the major gradient variables (longitude, latitude, altitude, TOC, TP, conductivity; the latter three log transformed). Variance partitioning by redundancy analysis (RDA), using function *varpart* in R package vegan (Oksanen *et al.*, 2013) on Hellinger transformed, normalized abundance data was used to estimate the variance fractions of eukaryote community composition that could be explained by the local environment, spatial gradients, or shared between them. The local environment was represented by concentrations of TOC, TP, and conductivity (all log transformed), while the spatial gradients were represented by longitude, latitude, and altitude. Univariate variance partitionings with the same predictor variables were done using LMs on richness and diversity indices, as well as NMDS site scores on two axes.

Spatial variance structures were investigated with the spdep package for R (Bivand & Piras, 2015). Spatial analysis depends on defining a neighbourhood relationship between sites, which unfortunately can be done many ways (Bivand *et al.*, 2008). We used two conceptually different neighbourhood definitions (Gabriel and Relative neighbour) which nevertheless generally produced similar results. Spatial patterns in model residuals were assessed with Moran's *I*

249 coefficient for spatial autocorrelation (Moran, 1950), using the *moran.test* function from the
250 spdep R package.

251 Relationships between functional traits and environment were investigated by so-called fourth
252 corner analysis (Legendre *et al.*, 1997) where the site by OTU matrix is pre- and post-multiplied
253 with matrices representing environment by site and OTU by trait. We used the same six
254 environmental variables in Fig. S2, and autotrophic, phagotrophic, or parasitic nutritional modes
255 as functional traits based on Simon *et al.* (2015). Significance tests used the combination of row-
256 and column-wise permutations of the site by OTU matrix (Dray & Legendre, 2008), as
257 implemented in the *fourthcorner* function of the ade4 package (Dray & Dufour, 2007).

258 Nestedness is usually defined as a biogeographical pattern where species-poor communities
259 form nested subsets of the richer ones. While the nestedness concept is old, quantitative methods
260 for detecting such patterns are more recent and still developing (Ulrich *et al.*, 2009). Nestedness
261 indices are usually tested against null model distributions generated by randomizations of the site
262 by OTU presence/absence matrix. Since we are basically interested in whether the distribution of
263 species across sites are random or not given that some species are common and some rare, we
264 used the RANDNEST algorithm of Jonsson (2001), which is a non-sequential randomization that
265 preserves OTU frequencies across sites but not row and column totals. Nestedness analysis was
266 performed with *oecosimu* function in the vegan package, using the discrepancy index of Brualdi
267 & Sanderson (1999) and the NODF index of Almeida‐Neto *et al.* (2008). We also used the
268 same vegan function to compute the relative beta diversity contributions from nestedness and
269 spatial turnover according to Baselga (2010). Presence/absence matrices were based on rarefied
270 reads using the same random seed as for calculating rarefied richness and diversity measures.
271

Results

Overall protistan community composition and taxonomic distribution

Non-protistan sequences amplified by the general eukaryotic primers (7.19 % of OTUs), mainly crustaceans and rotifers, were excluded. A total of 1882 OTUs were recovered from the 281 571 high-quality, denoised reads. An average of 426 (range: 145-771) OTUs were detected per sample and the mean number of total reads per lake was 3805 (range: 1107 - 6325). Protistan sequences were distributed across the supergroups Amoebozoa, Archaeplastida, Excavata, Opisthokonta, SAR (Stramenopiles, Alveolata and Rhizaria) as well as across several lineages of uncertain phylogenetic placement like Apusomonadida, Centroheliozoa, Cryptophyta, Haptophyta, Katablepharida and Telonemia (Fig. 2, Table S2). Although the relative abundance of taxonomic groups varied between lakes, the majority of sequences in all samples belong to Alveolata, Cryptophyta and Stramenopiles.

Alveolata was the most diverse supergroup (38.9 % of reads, 40.28% of OTUs). Among alveolates, dinoflagellates represented the biggest fraction in terms of reads (24.02%) and OTUs (20.40%). Ciliates (11.85%), Perkinsea (5.37%) and other alveolates (2.66%) constituted a significantly smaller proportion of OTUs. Interestingly, Cryptophyta sequences were consistently abundant, representing up to 26.60 % of the total reads, but not very diverse (7.12% of OTUs). Three OTUs affiliated to the Cryptomonadales family, were present in all lakes. Finally Stramenopiles was the third most abundant group in our study (18.27 % of reads), of which Chrysophyceae (11.10 % of reads) were present in all samples. Chrysophyceae comprised the second most diverse group (12.86 % of OTUs) after dinoflagellates.

Opisthokonts were dominated by fungi (6.64% of OTUs), followed by choanoflagellates (1.65% of OTUs) and ichthyosporeans (0.53% of OTUs). Chlorophyta, the most abundant

295 Archaeplastida group, were present in all lakes accounting for 3.26 % of reads and 7.12 % of
296 OTUs. Charophyta, the second most abundant archaeplastidan group, represents 0.19% of total
297 reads and 0.53% of OTUs. Haptophytes occurred at low proportions (0.20% of the total reads
298 and 0.43% of OTUs) and were assigned to freshwater Pavlovophyceae and Prymnesiophyceae or
299 undefined haptophyte sequences. Rhizaria (3.81 % of the total reads) were exclusively composed
300 of cercozoans (7.49% OTUs). Katablepharid reads comprised 1.28 % of the reads and 0.58 %
301 OTUs. A small proportion of the OTUs were assigned to Centroheliozoa (3.35 %), Excavata
302 (0.32 %), Amoebozoa (0.21 %), Telonemia (0.21 %) and Apusomonadida (0.05 %). Finally, four
303 OTUs (0.21 %) could not be assigned with confidence to any of the above mentioned groups.
304 Classical morphospecies taxa like Cryptophyta, Katablepharida and Synurophyceae were
305 "supradiverse", containing a larger than average fraction of the total reads and a lower than
306 average fraction of the OTUs (Fig. 2, Table S2). In contrast, largely heterotrophic groups such as
307 Fungi, Bicosoecida and Excavata, which are either hard to identify or neglected in classical
308 phytoplankton microscopy, were "superdiverse" maintaining higher than average OTU richness
309 despite lower than average read abundance (Fig. 2, Table S2).
310 The 20 most frequent OTUs, representing 38.40 % of reads (Fig. 3), were affiliated to
311 Cryptophyta, Stramenopiles (Chrysophyceae, Synurophyceae, Raphidophyceae), Alveolata
312 (Dinoflagellata), Centroheliozoa and Archaeplastida (Chlorophyta). The 1765 most infrequent
313 OTUs (93.80 % of all OTUs) represented only 34.30 % of reads.
314
315 *Richness across samples*
316 The 74 lake ecosystems differed in richness and diversity (Table S3). Rarefaction curves of OTU
317 richness (Fig. 4) for each lake indicated that the total eukaryotic diversity was not recovered in

318 any of the lakes. However, the overall SAC based on the progressive addition of samples show
319 that the gamma diversity in the studied area has been fully recovered (Fig. 4, insert). Shannon
320 diversity varied greatly across samples (range: 1.88 - 5.15) (Table S3) and OTU richness
321 increased towards the east (Fig. 5A).
322
323 *Environmental factors influencing eukaryotic community composition*
324 NMDS ordination axes based on Bray-Curtis distances from site-normalized read abundance
325 data were highly correlated with the corresponding DCA axes ($p < 0.0001$, Table S4). Similar
326 results by the two methods strongly suggest that a reliable gradient structure has been found.
327 Protistan community composition was significantly related to spatial and environmental
328 gradients ($p=0.001$ on 999 permutations). Vectors representing longitude and TOC (the latter log
329 transformed) pointed in the same direction (Fig. 5A), while being orthogonal to vectors
330 reflecting local environment (TP and specific conductivity; both log transformed) and shorter
331 spatial gradients (latitude and altitude). Community dissimilarity increased significantly with
332 geographical distance (Mantel correlation = 0.37, $p = 0.001$ on 999 permutations). Thus, we infer
333 that longitude is the strongest spatial factor influencing eukaryotic community composition in the
334 studied lakes.
335 Vectors representing GLM fits between NMDS axis scores and rarefied richnesses of the most
336 abundant taxonomic groups are shown in Fig. 5B, with the corresponding permutation-based
337 significance probabilities in Table S5. Total OTU richness was associated with both NMDS axes
338 which altogether explained 47% of the total deviance, while this fraction was generally lower for
339 individual taxonomic groups. 17 of the 25 groups were significantly related to NMDS scores at
340 an overall false discovery rate (FDR) <5%. Of these, Chrysophyceae and Dinoflagellata had the

closest associations with NMDS1, while Oomycota, Excavata, and Charophyta had the closest associations with NMDS2. The other groups were to a larger extent associated with both axes, and many of them closely aligned with the longitude vector (especially Raphidophyceae, Synurophyceae, Chytridiomycota, Choanozoa, and Cercozoa). OTU richnesses of Cryptophyta, Kathablepharida, Chlorophyta, and Bacillariophyceae were not significantly related to any ordination axis.

Functional groups' distribution

To investigate the distribution pattern of different eukaryotic groups, we classified the OTUs in four functional traits on the basis of their taxonomy and mode of nutrition: autotrophs, heterotrophs, parasites and unclassified (Adl *et al.*, 2012, Simon *et al.*, 2015). Some of these groups are considered obligate autotrophs (e.g. diatoms, chlorophytes, charophytes, dictyochophytes, raphidophytes and synurophytes) while others are putative mixotrophs (e.g. chrysophytes, cryptophytes and haptophytes) (Jones, 2000, Adl *et al.*, 2012). Heterotrophic protists were represented by bicosoecids, centroheliozoans, choanozoans, ciliates and saprotrophic fungi. Putative parasites included members of Cercozoa, Chytridiomycota (Gleason *et al.*, 2008), Cryptomycota (Gleason *et al.*, 2012), Ichthyosporea (Glockling *et al.*, 2013) and Perkinsea (Bråte *et al.*, 2010, Mangot *et al.*, 2011). The fourth corner analysis showed no significant associations between functional traits and any of the indicator variables for spatial position or local environment (all FDR-corrected p-values > 0.34 on 999 permutations). Visual inspection of distribution patterns of functional traits (Fig. S4) supported the same conclusion. However, performing the fourth corner analysis with mixotrophy as an additional trait (i.e. identifying cryptophytes, chrysophytes and haptophytes as putative mixotrophs) indicated

364 significant relationships between longitude and both autotrophy and mixotrophy (adjusted p-
365 values on 999 permutations = 0.0315 and 0.0138, respectively). These results demonstrate that
366 distinguishing between autotrophs and mixotrophs enhanced the resolution of our analyses.
367 Autotrophs and mixotrophs comprised 53.6% of the OTUs, while 19.4% of the heterotrophic
368 OTUs belonged to putative parasitic groups. In terms of relative abundance, autotrophs
369 dominated (71.21% total reads), with heterotrophs and parasites constituting 16.61% and 8.30%
370 of all reads, respectively. In other words, heterotrophic and parasitic groups were represented by
371 fewer reads per OTU than primary producers.
372 Mean observed richness of autotrophs (and mixotrophs), heterotrophs and parasites were 249
373 (range = 83-454; SD = 68), 85 (range = 20-157; SD = 29), and 61 (range = 14-112; SD = 22)
374 OTUs per lake, respectively. The corresponding beta diversities (= total number of OTUs /
375 average number of OTUs per lake; often interpreted as the number of community turnovers
376 along the main gradient), were 4.1, 4.2, and 6.0 for autotrophs (and mixotrophs), heterotrophs
377 and parasites, respectively.
378
379 *Variance partitioning and spatial autocorrelation*
380 Variance partitioning by RDA on Hellinger transformed relative read abundances was used to
381 assess the contributions of environmental and spatial gradients to the compositional variation.
382 Both environmental factors and spatial gradients were significant in explaining parts of the
383 protistan community composition (5.2 and 3.6% adjusted of total variation, respectively, not
384 including the shared effect between local and spatial components). Approximately 90% of the
385 community variance remained unexplained by environmental and spatial gradient indicators
386 (Table S6). To further investigate the contribution of environmental vs. spatial predictors, we

chose alpha diversity measures (rarefied OTU richness, Shannon, Simpson) and NMDS site scores (k=2) as dependent variables using the same spatial and environmental factors as predictor variables. Rarefied OTU richness, Shannon diversity and Simpson diversity were closely correlated (Spearman's $\rho = 0.86 – 0.96$, Kendall's $\tau = 0.69 – 0.86$), indicating they capture the same aspects of diversity. Local environment and spatial factors explained 8% and 33% of OTU richness variation independently and 13% in combination (Table S6). Spatial factors alone explained 42% and 40% of total variation in Shannon and Simpson diversity in our lakes, respectively ($p < 0.001$) (Table S6). NMDS1 axis has stronger effects of local environment (33% of variance explained, Fig. S5, Table S6). In contrast, NMDS2 axis has variance partitioning pattern with slight dominance of both local and spatial predictors in comparison to spatial parameters only (15 and 13% of the variance explained, respectively) (Table S6). The Moran's *I* test (Moran, 1950) showed no residual spatial autocorrelation in protistan community richness (rarefied OTU richness, Shannon and Simpson diversity) beyond what could be explained by the spatial predictors (longitude, latitude, altitude). Both parametric and resampling-based probabilities were non-significant.

Nestedness analysis

Both the tested nestedness indicators (BS-discrepancy and NODF) showed highly significant results ($p = 0.01$ on 99 randomizations), with standardized effect sizes of -25 and 78, respectively. Partitioning beta diversity according to Baselga (2010) indicated an overwhelming (> 99%) contribution from spatial turnover. In other words, while nestedness was a statistically significant feature of our dataset, it accounted for less than 1% of the total beta diversity, irrespective of whether the analysis was based on Jaccard and Sørensen indices.

Discussion

Microbial diversity in freshwater habitats is still poorly described and under-sampled (Lefèvre *et al.*, 2008), although an increasing number of HTS surveys is improving the situation. Many freshwater HTS studies reported so far have been seasonal studies in single locations (Lepère *et al.*, 2013, Mangot *et al.*, 2013) or from a small number of lakes with special properties like high salinity or altitude (Wang *et al.*, 2014, Kammerlander *et al.*, 2015). Our study is the first to cover an extensive spatial gradient on a set of lowland lakes carefully selected to be as homogeneous as possible with respect to factors other than local productivity and spatial position. By choosing a synoptic sampling strategy we deliberately emphasize the comparability between lakes at the expense within-lake representability of a single snapshot sample. Despite its limitations, our HTS study clearly confirms of the longitudinal biodiversity gradient in Southern Scandinavia that has earlier been inferred from microscopy (Hessen *et al.*, 2006, Ptacnik *et al.*, 2010).

Spatial factors (longitude, latitude, altitude) explained a higher fraction of the protistan community variation than environmental factors (TOC, TP and conductivity), although a substantial fraction could not be resolved between the two sources of variation. A larger fraction of total variance (40%) was explainable in LMs of OTU richness than in RDA models of community composition (9%). The level of explained RDA variance is comparable to some microscopy-based phytoplankton community studies (Beisner *et al.*, 2006), while others have reported higher fractions. For example, the meta-study of lentic diatom communities in Eurasia, Africa and Antarctica by Verleyen *et al.* (2009) found environmental factors to account for most of the community variation (21%) while spatial factors related to dispersal (5.5%) were less important.

433 Longitude had stronger influence on the eukaryotic community composition in Scandinavian
434 lakes than other spatial variables, as observed by Stomp *et al.* (2011) in North American lakes.
435 By constraining lake size and productivity in our sampling design, we think that our study has
436 less chance of being confounded by large-scale patterns in lake shape and local productivity
437 compared to Stomp *et al.* (2011). Overall, surveys conducted in lakes point to the significance of
438 both spatial and environmental predictors as drivers of community structure, with the relative
439 effect of these factors is likely dependent on environmental gradients, spatial extent and dispersal
440 ability (Heino *et al.*, 2015).
441
442 *Plankton community composition patterns*
443 Microscopy-based phytoplankton surveys (Watson *et al.*, 1997, Ptacnik *et al.*, 2008a) have found
444 that the relative biomass contributions of chrysophytes and cryptophytes decrease while diatoms,
445 chlorophytes, and cyanobacteria increase with increasing eutrophication (for which TP is the
446 most common proxy). The high read abundances of chrysophytes and cryptophytes in our study
447 concurs with previously reported pattern. Chrysophyceae dominance has been frequently related
448 to the extreme oligotrophic conditions, and their disappearance indicates eutrophication (Ptacnik
449 *et al.*, 2008a). Supplementing nutrient uptake by phagotrophy of bacteria and small
450 phytoplankton may be an important adaptation to nutrient-poor conditions in chrysophytes
451 (Jones, 2000).
452 Watson *et al.* (1997) report relative biomass contributions of dinoflagellates half that of
453 chrysophytes and cryptophytes, which is somewhat in contrast with the dominance by alveolate
454 and dinoflagellate OTUs in our study. This discrepancy may partly be due to underestimation by
455 fixation losses and identification problems for especially non-thecate dinoflagellates by

456 microscopy, and partly overestimation due to biases in molecular methods. Alveolates are known
457 to have high rDNA copy numbers (Zhu *et al.*, 2005, Massana, 2011, Gong *et al.*, 2013) which
458 could lead to overestimation of cell abundance, but not necessarily of biomass. The well-
459 documented relationship between cell size and genome size across all eukaryotes (Prokopowich
460 *et al.*, 2003) implies that OTUs with high relative read abundance due to high rDNA copy
461 number would also be expected to have high contribution to total biomass, as has been recently
462 shown for haptophytes (Egge *et al.*, 2013). While rRNA genes in other protists are generally
463 thought to evolve in a strictly conserved manner such that all repeats will be identical (Dover,
464 1982), this seems to not always the case for ciliates (Gong *et al.*, 2013). Hence, some of the
465 diversity of alveolate rDNA could be attributed to intra-genomic variability that is especially
466 valid for larger cells with high number of rDNA copies.
467 Fungi (saprotrophs and parasites) were less diverse in our study (6.64% OTUs) than in other
468 recent reports (Lefranc *et al.*, 2005, Lepère *et al.*, 2008, Mangot *et al.*, 2013). Chytridiomycota
469 are commonly found on a large variety of substrates in freshwater and decompose chitin, starch
470 and cellulose in detrital organic materials (Gleason *et al.*, 2008), such that the observed positive
471 association with TOC could reflect direct utilization of dissolved organic matter. Conversely,
472 chytrids are also commonly described as phytoplankton parasites that respond chemotactically to
473 photosynthesis exudates, and frequently parasitize diatoms (Kagami *et al.*, 2012, Kagami *et al.*,
474 2015), which is supported by their proximity in the NMDS ordination. The longitudinal
475 alignment and eastward increase in Raphidophyceae richness is consistent with indications of
476 ongoing dispersal by the nuisance raphidophyte *Gonyostomum semen* towards western
477 Scandinavia (Hongve *et al.*, 1988, Lepistö *et al.*, 1994, Rengefors *et al.*, 2012).
478

Functional diversity and nestedness

While there were no significant relationships between spatial and environmental gradients and lumped autotrophic, heterotrophic or parasitic traits (Fig. S4), significant fourth corner relationships were found when the autotrophic trait was differentiated into pure photoautotrophy and mixotrophy. Since both traits were associated with longitude, we infer a longitudinal shift with increasing mixotrophy from west to east. In contrast, heterotrophic and parasitic lifestyles all appear to have essential roles in the organization, energy transfer, and element cycling of aquatic food webs (Gleason *et al.*, 2008, Rasconi *et al.*, 2011, Rasconi *et al.*, 2012). Beta diversity (Tuomisto, 2010) for primary producers and heterotrophs was lower than for parasites indicating a higher species packing and therefore narrower ecological niche widths for parasitic taxa.

The classical nestedness analysis methods have all been developed on pre-HTS data sets with typically at least an order of magnitude fewer species than the number of OTUs in the present study. Fayle & Manica (2010) found that inflated Type 1 error rates were a general problem for HTS-type data sets, and that all common permutation strategies were affected by this problem. In our data set, the contribution of nestedness to total beta diversity appeared to be very minor even though tests for nestedness per se were highly significant. We think this illustrates a classical conflict between statistical and biological significance in large data sets (e.g. Yoccoz (1991)), and conclude that spatial turnover is probably the most important source of beta diversity in Nordic lakes.

Concluding remarks

While improved sampling strategies or increased sequencing depth most likely would have given a more complete picture of local (alpha) diversity, we are confident that the over-all pattern in diversity and community composition is well captured by our study. Our analyses indicated that spatial position is more important than the local environment for the composition and diversity of protistan communities in Northern lakes. Altitude and latitude are, in contrast to longitude, established proxies for well-documented climate effects on biodiversity (Gaston, 2000). We deliberately constrained the climatic variation in our study by making the longitudinal gradient three times longer than the latitudinal. With this design we find a strong longitudinal signal of the same magnitude as in earlier studies with non-molecular methods (Hessen *et al.*, 2006, Ptacnik *et al.*, 2010).

The mechanistic explanation for the longitudinal biodiversity gradient in Scandinavia has been elusive, mainly because it is aligned with geographical gradients in climate, soil depth, terrain ruggedness and landscape productivity. While climatic gradients have been designed to be as short as possible in our study, they can only be entirely eliminated at the expense of spatial extent. The study area has longitudinal oceanicity gradient with an eastward increase in the difference between winter and summer temperatures, while there is no longitudinal trend in mean annual temperature (Fig. S6). We nevertheless feel that the longitudinal biodiversity gradient in Scandinavia cannot just be dismissed as a climatic gradient in disguise.

By being open, interactive habitat patches, lake ecosystems are often considered epitomes of the meta-community concept, where the opposing forces of dispersal and environmental filtering are the main biodiversity-shaping forces (Leibold *et al.*, 2004). The relative importance of dispersal in lentic plankton communities is difficult to quantify directly, but is expected to depend on both the scale and heterogeneity of the landscape in which the lakes are embedded.

Data are limited, but reviews by Shurin *et al.* (2009) and Jenkins (2014) indicate that dispersal limitation is relevant for lake biota, but perhaps less so for planktonic protists than for larger organisms. Ptacnik *et al.* (2010) argue that since dissimilarity between planktonic communities tend to increase with productivity, landscapes where productive lakes are common will tend to maintain larger regional species pools than poorer ones. Regional differences in landscape productivity are reflected by the regionally averaged TP (TPreg) of Ptacnik *et al.* (2010), which is closely related to longitude in our study area (Fig. S1). In this perspective, the longitudinal biodiversity gradient in Scandinavia can be seen as the contrast between species-rich lakes in the productive landscapes of the east with poorer landscapes where productive lakes are rarer in the west.

Acknowledgements

This study has been supported financially by the Department of Biosciences, University of Oslo and by the Research Council of Norway (contract Miljø2015/196336 "Biodiversity, community saturation and ecosystem function in lakes" (COMSAT)). We thank the COMSAT field sampling crew, especially Dag. O. Hessen, Johnny Håll, Marcia Kyle, Robert Ptacnik, and Jan-Erik Thrane, for their efforts. Robert Ptacnik is acknowledged for providing the TPreg data.

Conflict of Interest

The authors declare no conflict of interest.

References

Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M & Hampl V (2012) The revised classification of eukaryotes. *J Euk Microbiol* **59**: 429-514.

Almeida‐Neto M, Guimaraes P, Guimarães PR, Loyola RD & Ulrich W (2008) A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**: 1227-1239.

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Aylward B, Bandyopadhyay J, Belausteguigotia J-C, Borkey P, Cassar A, Meadors L, Saade L, Siebentritt M, Stein R & Tognetti S (2005) Freshwater ecosystem services. Vol. 3 (Chopra KR, ed.) p.^pp. 213-256. Island Press.

Balzano S, Abs E & Leterme SC (2015) Protist diversity along a salinity gradient in a coastal lagoon. *Aquat Microb Ecol* **74**: 263-277.

Baselga A (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* **19**: 134-143.

Bates ST, Clemente JC, Flores GE, Walters WA, Parfrey LW, Knight R & Fierer N (2013) Global biogeography of highly diverse protistan communities in soil. *The ISME journal* **7**: 652-659.

Beisner BE, Peres-Neto PR, Lindström ES, Barnett A & Longhi ML (2006) The role of environmental and spatial processes in structuring lake communities from bacteria to fish. *Ecology* **87**: 2985-2991.

Bivand R & Piras G (2015) Comparing implementations of estimation methods for spatial econometrics. *J Stat Softw* **63**: 1-36.

Bivand RS, Pebesma EJ, Gomez-Rubio V & Pebesma EJ (2008) *Applied spatial data analysis with R*. Springer, NY.

Brettum P (1989) *Alger som indikator på vannkvalitet. Planteplankton*. Norsk institutt for vannforskning.

Brettum P & Andersen T (2004) The use of phytoplankton as indicators of water quality. *Norwegian Institute for Water Research SNO Report* 33.

Brualdi RA & Sanderson JG (1999) Nested species subsets, gaps, and discrepancy. *Oecologia* **119**: 256-264.

Bråte J, Logares R, Berney C, Ree DK, Klaveness D, Jakobsen KS & Shalchian-Tabrizi K (2010) Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *ISME J* **4**: 1144-1153.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK & Gordon JI (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335-336.

Carlsen T, Aas AB, Lindner D, Vrålstad T, Schumacher T & Kauserud H (2012) Don't make a mista (g) ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecol* **5**: 747-749.

Charvet S, Vincent WF, Comeau A & Lovejoy C (2012) Pyrosequencing analysis of the protist communities in a High Arctic meromictic lake: DNA preservation and change. *Front Microbiol* **3**: 1-14.

Cole JJ, Prairie YT, Caraco NF, McDowell WH, Tranvik LJ, Striegl RG, Duarte CM, Kortelainen P, Downing JA & Middelburg JJ (2007) Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. *Ecosystems* **10**: 172-185.

De Meester L, Declerck S, Stoks R, Louette G, Van De Meutter F, De Bie T, Michels E & Brendonck L (2005) Ponds and pools as model systems in conservation biology, ecology and evolutionary biology. *Aquat Conserv* **15**: 715-725.

de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N & Probert I (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**: 1261605.

Doherty M, Costas BA, McManus GB & Katz LA (2007) Culture-independent assessment of planktonic ciliate diversity in coastal northwest Atlantic waters. *Aquat Microb Ecol* **48**: 141-154.

592 Dover G (1982) Molecular drive: a cohesive mode of species evolution. *Nature* **299**: 111-117.
593 Dray S & Dufour A-B (2007) The ade4 package: implementing the duality diagram for ecologists. *J Stat*
594 *Softw* **22**: 1-20.
595 Dray S & Legendre P (2008) Testing the species traits-environment relationships: the fourth-corner
596 problem revisited. *Ecology* **89**: 3400-3412.
597 Dudgeon D, Arthington AH, Gessner MO, Kawabata ZI, Knowler DJ, Lévêque C, Naiman RJ, Prieur‐
598 Richard AH, Soto D & Stiassny ML (2006) Freshwater biodiversity: importance, threats, status and
599 conservation challenges. *Biol Rev* **81**: 163-182.
600 Dunthorn M, Klier J, Bunge J & Stoeck T (2012) Comparing the Hyper‐Variable V4 and V9 Regions of the
601 Small Subunit rDNA for Assessment of Ciliate Environmental Diversity. *J Eukaryot Microbiol* **59**: 185-187.
602 Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-
603 2461.
604 Egge E, Bittner L, Andersen T, Audic S, de Vargas C & Edvardsen B (2013) 454 pyrosequencing to describe
605 microbial eukaryotic community composition, diversity and relative abundance: a test for marine
606 haptophytes. *PLoS ONE* **8**: e74371.
607 Elser JJ, Andersen T, Baron JS, Bergström A-K, Jansson M, Kyle M, Nydick KR, Steger L & Hessen DO (2009)
608 Shifts in lake N: P stoichiometry and nutrient limitation driven by atmospheric nitrogen deposition.
609 *science* **326**: 835-837.
610 Fayle TM & Manica A (2010) Reducing over-reporting of deterministic co-occurrence patterns in biotic
611 communities. *Ecological Modelling* **221**: 2237-2242.
612 Gaston KJ (2000) Global patterns in biodiversity. *Nature* **405**: 220-227.
613 Gleason FH, Kagami M, Lefevre E & Sime-Ngando T (2008) The ecology of chytrids in aquatic ecosystems:
614 roles in food web dynamics. *Fungal Biol Rev* **22**: 17-25.
615 Gleason FH, Carney LT, Lilje O & Glockling SL (2012) Ecological potentials of species of Rozella
616 (Cryptomycota). *Fungal Ecol* **5**: 651-656.
617 Glockling SL, Marshall WL & Gleason FH (2013) Phylogenetic interpretations and ecological potentials of
618 the Mesomycetozoea (Ichthyosporea). *Fungal Ecol* **6**: 237-247.
619 Gong J, Dong J, Liu X & Massana R (2013) Extremely high copy numbers and polymorphisms of the rDNA
620 operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Protist* **164**: 369-379.
621 Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, De Vargas C & Decelle J
622 (2012) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit
623 rRNA sequences with curated taxonomy. *Nucleic Acids Res* 1-8.
624 Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK &
625 Sodergren E (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-
626 pyrosequenced PCR amplicons. *Genome Res* **21**: 494-504.
627 Heino J (2013) The importance of metacommunity ecology for environmental assessment research in
628 the freshwater realm. *Biol Rev* **88**: 166-178.
629 Heino J, Melo AS, Siqueira T, Soininen J, Valanko S & Bini LM (2015) Metacommunity organisation,
630 spatial extent and dispersal in aquatic systems: patterns, processes and prospects. *Freshw Biol* **60**: 845-
631 869.
632 Heino J, Bini LM, Karjalainen SM, Mykrä H, Soininen J, Vieira LCG & Diniz‐Filho JAF (2010) Geographical
633 patterns of micro‐organismal community structure: are diatoms ubiquitously distributed across boreal
634 streams? *Oikos* **119**: 129-137.
635 Henriksen A, Skjelvåle BL, Mannio J, Wilander A, Harriman R, Curtis C, Jensen JP, Fjeld E & Moiseenko T
636 (1998) Northern European lake survey, 1995: Finland, Norway, Sweden, Denmark, Russian Kola, Russian
637 Karelia, Scotland and Wales. *Ambio* 80-91.
638 Hessen DO, Faafeng BA, Smith VH, Bakkestuen V & Walseng B (2006) Extrinsic and intrinsic controls of
639 zooplankton diversity in lakes. *Ecology* **87**: 433-443.

Hessen DO, Andersen T, Larsen S, Skjelkvåle BL & de Wit HA (2009) Nitrogen deposition, catchment productivity, and climate as determinants of lake stoichiometry. *Limnol Oceanogr* **54**: 2520-2528.
Hijmans RJ, Cameron SE, Parra JL, Jones PG & Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* **25**: 1965-1978.
Hill MO & Gauch Jr HG (1980) Detrended correspondence analysis: an improved ordination technique. *Vegetatio* **42**: 47-58.
Hongve D, Lovstad O & Bjorndalen K (1988) Gonyostomum semen - a new nuisance to bathers in Norwegian lakes. *Verh Internat Verein Limnol* **23**: 430-434.
Hortal J, Nabout JC, Calatayud J, Carneiro FM, Padial A, Santos A, Siqueira T, Bokma F, Bini LM & Ventura M (2014) Perspectives on the use of lakes and ponds as model systems for macroecological research. *J Limnol* **73**: 46-60.
Jansson M, Bergström A-K, Blomqvist P & Drakare S (2000) Allochthonous organic carbon and phytoplankton/bacterioplankton production relationships in lakes. *Ecology* **81**: 3250-3255.
Jenkins DG (2014) Lakes and rivers as microcosms, version 2.0. *Journal of Limnology* **73**.
Johnson ZI & Martiny AC (2015) Techniques for Quantifying Phytoplankton Biodiversity. *Ann Rev Mar Sci* **7**: 299-324.
Jones RI (2000) Mixotrophy in planktonic protists: an overview. *Freshw Biol* **45**: 219-226.
Jonsson BG (2001) A null model for randomization tests of nestedness in species assemblages. *Oecologia* **127**: 309-313.
Kagami M, Amano Y & Ishii N (2012) Community structure of planktonic fungi and the impact of parasitic chytrids on phytoplankton in Lake Inba, Japan. *Microbial ecology* **63**: 358-368.
Kagami M, Miki T & Takimoto G (2015) Mycoloop: chytrids in aquatic food webs. *Roles and mechanisms of parasitism in aquatic microbial communities* 36.
Kammerlander B, Breiner H-W, Filker S, Sommaruga R, Sonntag B & Stoeck T (2015) High diversity of protistan plankton communities in remote high mountain lakes in the European Alps and the Himalayan mountains. *FEMS Microbiol Ecol* **91**: fiv010.
Kontula T & Väinölä R (2001) Postglacial colonization of Northern Europe by distinct phylogeographic lineages of the bullhead, Cottus gobio. *Mol Ecol* **10**: 1983-2002.
Kunin V, Engelbrektson A, Ochman H & Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12**: 118-123.
Lefèvre E, Roussel B, Amblard C & Sime-Ngando T (2008) The molecular diversity of freshwater picoeukaryotes reveals high occurrence of putative parasitoids in the plankton. *PLoS ONE* **3**: e2324.
Lefranc M, Thénot A, Lepere C & Debroas D (2005) Genetic diversity of small eukaryotes in lakes differing by their trophic status. *Appl Environ Microbiol* **71**: 5935-5942.
Legendre P, Galzin R & Harmelin-Vivien ML (1997) Relating behavior to habitat: Solutions to thefourth-corner problem. *Ecology* **78**: 547-562.
Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase J, Hoopes M, Holt R, Shurin J, Law R & Tilman D (2004) The metacommunity concept: a framework for multi‐scale community ecology. *Ecol Lett* **7**: 601-613.
Lepère C, Domaizon I & Debroas D (2008) Unexpected importance of potential parasites in the composition of the freshwater small-eukaryote community. *Appl Environ Microbiol* **74**: 2940-2949.
Lepère C, Boucher D, Jardillier L, Domaizon I & Debroas D (2006) Succession and regulation factors of small eukaryote community composition in a lacustrine ecosystem (Lake Pavin). *Appl Environ Microbiol* **72**: 2971-2981.
Lepère C, Domaizon I, Taïb N, Mangot J-F, Bronner G, Boucher D & Debroas D (2013) Geographic distance and ecosystem size determine the distribution of smallest protists in lacustrine ecosystems. *FEMS Microbiol Ecol* **85**: 85-94.

688 Lepistö L, Antikainen S & Kivinen J (1994) The occurrence of Gonyostomum semen (Ehr.) Diesing in
689 Finnish lakes. *Hydrobiologia* **273**: 1-8.
690 Logares R, Audic S, Santini S, Pernice MC, de Vargas C & Massana R (2012) Diversity patterns and activity
691 of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *ISME J* **6**: 1823-1833.
692 Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R, Claverie J-M, Decelle J, Dolan JR & Dunthorn
693 M (2014) Patterns of rare and abundant marine microbial eukaryotes. *Curr Biol* **24**: 813-821.
694 Mangot J-F, Debroas D & Domaizon I (2011) Perkinsozoa, a well-known marine protozoan flagellate
695 parasite group, newly identified in lacustrine systems: a review. *Hydrobiologia* **659**: 37-48.
696 Mangot JF, Domaizon I, Taib N, Marouni N, Duffaud E, Bronner G & Debroas D (2013) Short‐term
697 dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes.
698 *Environ Microbiol* **15**: 1745-1758.
699 Martiny JBH, Bohannan BJ, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M,
700 Krumins JA & Kuske CR (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev*
701 *Microbiol* **4**: 102-112.
702 Massana R (2011) Eukaryotic picoplankton in surface oceans. *Annu Rev Microbiol* **65**: 91-110.
703 Massana R, del Campo J, Sieracki ME, Audic S & Logares R (2014) Exploring the uncultured
704 microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *ISME J* **8**: 854-
705 866.
706 Minchin PR (1987) An evaluation of the relative robustness of techniques for ecological ordination.
707 *Vegetatio* **69**: 89-107.
708 Moran PA (1950) Notes on continuous stochastic phenomena. *Biometrika* 17-23.
709 Nolte V, Pandey RV, Jost S, Medinger R, Ottenwaelder B, Boenigk J & Schloetterer C (2010) Contrasting
710 seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Mol*
711 *Ecol* **19**: 2908-2915.
712 O'Neil J, Davis TW, Burford MA & Gobler C (2012) The rise of harmful cyanobacteria blooms: the
713 potential roles of eutrophication and climate change. *Harmful Algae* **14**: 313-334.
714 Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara R, Simpson GL, Solymos P, Stevens
715 MHH & Wagner H (2013) vegan: Community Ecology Package. R package version 2.3-2.
716 Padisak J, Borics G, Grigorszky I & Soroczki-Pinter E (2006) Use of phytoplankton assemblages for
717 monitoring ecological status of lakes within the Water Framework Directive: the assemblage index.
718 *Hydrobiologia* **553**: 1-14.
719 Pianka ER (1974) *Evolutionary Ecology*. New York, USA.
720 Prokopowich CD, Gregory TR & Crease TJ (2003) The correlation between rDNA copy number and
721 genome size in eukaryotes. *Genome* **46**: 48-50.
722 Ptacnik R, Andersen T, Brettum P, Lepistö L & Willén E (2010) Regional species pools control community
723 saturation in lake phytoplankton. *Proc Biol Sci* **277**: 3755-3764.
724 Ptacnik R, Lepistö L, Willén E, Brettum P, Andersen T, Rekolainen S, Solheim AL & Carvalho L (2008a)
725 Quantitative responses of lake phytoplankton to eutrophication in Northern Europe. *Aquat Ecol* **42**: 227-
726 236.
727 Ptacnik R, Solimini AG, Andersen T, Tamminen T, Brettum P, Lepistö L, Willén E & Rekolainen S (2008b)
728 Diversity predicts stability and resource use efficiency in natural phytoplankton communities. *Proc Natl*
729 *Acad Sci USA* **105**: 5134-5138.
730 Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J & Glöckner FO (2012) The SILVA
731 ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*
732 **41**: gks1219.
733 Quince C, Lanzén A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF & Sloan WT (2009) Accurate
734 determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* **6**: 639-641.

735 R Development Core Team (2014) R: A Language and Environment for Statistical Computing. R
736 Foundation for Statistical Computing, Vienna, Austria.
737 Rasconi S, Jobard M & Sime-Ngando T (2011) Parasitic fungi of phytoplankton: ecological roles and
738 implications for microbial food webs. *Aquat Microb Ecol* **62**: 123-137.
739 Rasconi S, Niquil N & Sime‐Ngando T (2012) Phytoplankton chytridiomycosis: community structure and
740 infectivity of fungal parasites in aquatic ecosystems. *Environ Microbiol* **14**: 2151-2170.
741 Reeder J & Knight R (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank-
742 abundance distributions. *Nat Methods* **7**: 668-669.
743 Refseth U, Nesbø C, Stacy J, Vøllestad L & Fjeld E (1998) Genetic evidence for different migration routes
744 of freshwater fish into Norway revealed by analysis of current perch (Perca fluviatilis) populations in
745 Scandinavia. *Mol Ecol* **7**: 1015-1027.
746 Rengefors K, Weyhenmeyer GA & Bloch I (2012) Temperature as a driver for the expansion of the
747 microalga Gonyostomum semen in Swedish lakes. *Harmful algae* **18**: 65-73.
748 Rosén G (1981) Tusen sjöar. *Växtplanktons miljökrav Liber förlag, Stockholm*.
749 Ryder R (1982) The morphoedaphic index—use, abuse, and fundamental concepts. *Trans Am Fish Soc*
750 **111**: 154-164.
751 Santoferrara LF, Grattepanche JD, Katz LA & McManus GB (2014) Pyrosequencing for assessing diversity
752 of eukaryotic microbes: analysis of data on marine planktonic ciliates and comparison with traditional
753 methods. *Environ Microbiol* **16**: 2752-2763.
754 Schindler D (1977) Evolution of phosphorus limitation in lakes. *Science* **195**: 260-262.
755 Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH
756 & Robinson CJ (2009) Introducing mothur: open-source, platform-independent, community-supported
757 software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.
758 Shurin JB, Cottenie K & Hillebrand H (2009) Spatial autocorrelation and dispersal limitation in freshwater
759 organisms. *Oecologia* **159**: 151-159.
760 Simon M, López-García P, Deschamps P, Moreira D, Restoux G, Bertolino P & Jardillier L (2015) Marked
761 seasonality and high spatial variability of protist communities in shallow freshwater systems. *ISME J* 1-13.
762 Šlapeta J, Moreira D & López-García P (2005) The extent of protist diversity: insights from molecular
763 ecology of freshwater eukaryotes. *Proc Biol Sci* **272**: 2073-2081.
764 Soininen J (2012) Macroecology of unicellular organisms–patterns and processes. *Environ Microbiol Rep*
765 **4**: 10-22.
766 Soininen J, Korhonen JJ, Karhu J & Vetterli A (2011) Disentangling the spatial patterns in community
767 composition of prokaryotic and eukaryotic lake plankton. *Limnology and Oceanography* **56**: 508-520.
768 Solheim AL, Rekolainen S, Moe SJ, Carvalho L, Phillips G, Ptacnik R, Penning WE, Toth LG, O'Toole C &
769 Schartau A-KL (2008) Ecological threshold responses in European lakes and their applicability for the
770 Water Framework Directive (WFD) implementation: synthesis of lakes results from the REBECCA project.
771 *Aquat Ecol* **42**: 317-334.
772 Stoeck T, Bass D, Nebel M, Christen R, Jones MD, Breiner HW & Richards TA (2010) Multiple marker
773 parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine
774 anoxic water. *Mol Ecol* **19**: 21-31.
775 Stomp M, Huisman J, Mittelbach GG, Litchman E & Klausmeier CA (2011) Large-scale biodiversity
776 patterns in freshwater phytoplankton. *Ecology* **92**: 2096-2107.
777 Strayer DL & Dudgeon D (2010) Freshwater biodiversity conservation: recent progress and future
778 challenges. *J North Am Benthol Soc* **29**: 344-358.
779 Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G &
780 Kõljalg U (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar
781 results but reveal substantial methodological biases. *New Phytol* **188**: 291-301.

782 Thrane J-E, Hessen DO & Andersen T (2014) The Absorption of Light in Lakes: Negative Impact of
783 Dissolved Organic Carbon on Primary Productivity. *Ecosystems* **17**: 1040-1052.
784 Triadó‐Margarit X & Casamayor EO (2012) Genetic diversity of planktonic eukaryotes in high mountain
785 lakes (Central Pyrenees, Spain). *Environ Microbiol* **14**: 2445-2456.
786 Tuomisto H (2010) A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining
787 beta diversity as a function of alpha and gamma diversity. *Ecography* **33**: 2-22.
788 Ulrich W, Almeida-Neto M & Gotelli NJ (2009) A consumer's guide to nestedness analysis. *Oikos* **118**: 3-
789 17.
790 Vainio JK & Väinölä R (2003) Refugial races and postglacial colonization history of the freshwater
791 amphipod Gammarus lacustris in Northern Europe. *Biol J Linn Soc Lond* **79**: 523-542.
792 Venables W & Ripley B (2002) *Modern Applied Statistics with S*. Springer Science, New York.
793 Verleyen E, Vyverman W, Sterken M, Hodgson DA, De Wever A, Juggins S, Van de Vijver B, Jones VJ,
794 Vanormelingen P & Roberts D (2009) The importance of dispersal related and local factors in shaping the
795 taxonomic structure of diatom metacommunities. *Oikos* **118**: 1239-1249.
796 Wang J, Wang F, Chu L, Wang H, Zhong Z, Liu Z, Gao J & Duan H (2014) High Genetic Diversity and
797 Novelty in Eukaryotic Plankton Assemblages Inhabiting Saline Lakes in the Qaidam Basin. *PLoS ONE* **9**:
798 e112812.
799 Watson SB, McCauley E & Downing JA (1997) Patterns in phytoplankton taxonomic composition across
800 temperate lakes of differing nutrient status. *Limnol Oceanogr* **42**: 487-495.
801 Willén E (2007) Växtplankton i sjöar.
802 Worden AZ (2006) Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquat Microb Ecol* **43**:
803 165-175.
804 Yoccoz NG (1991) Use, overuse, and misuse of significance tests in evolutionary biology and ecology.
805 *Bulletin of the Ecological Society of America* **72**: 106-111.
806 Zhao B, Chen M, Sun Y, Yang J & Chen F (2011) Genetic diversity of picoeukaryotes in eight lakes
807 differing in trophic status. *Can J Microbiol* **57**: 115-126.
808 Zhu F, Massana R, Not F, Marie D & Vaulot D (2005) Mapping of picoeucaryotes in marine ecosystems
809 with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol Ecol* **52**: 79-92.

810

811 **Figure Legends**

812 **Fig. 1.** A map of sampled lakes (n=77) colored by annual mean air temperature (ºC) based on the
813 WorldClim database (Hijmans *et al.*, 2005). Contour lines represent altitude below 600 m. The
814 mountain ridge extends S-N around 8ºE. Three lakes indicated by black color were not included
815 in the analysis due to glacial influence or high/low pH.

816

817 **Fig. 2.** Distribution of OTUs and reads (log-scale) of all detected groups and lineages of
818 unresolved phylogenetic placement. Each point represents a group/lineage, and point color
819 reflects taxonomic affiliation (supergroup): yellow = Archaeplastida, red = Alveolata, green =
820 Opisthokonta, blue = Stramenopiles.

821

822 **Fig. 3.** Rank-abundance curve for the total 1882 OTUs detected in 74 lakes. The relative
823 abundance of top 20 protistan OTUs is shown in the insert. The identity number of the respective
824 OTU is shown below the bars. Colors represent the supergroups: yellow = Archaeplastida, red =
825 Alveolata, green = Opisthokonta, blue = Stramenopiles.

826

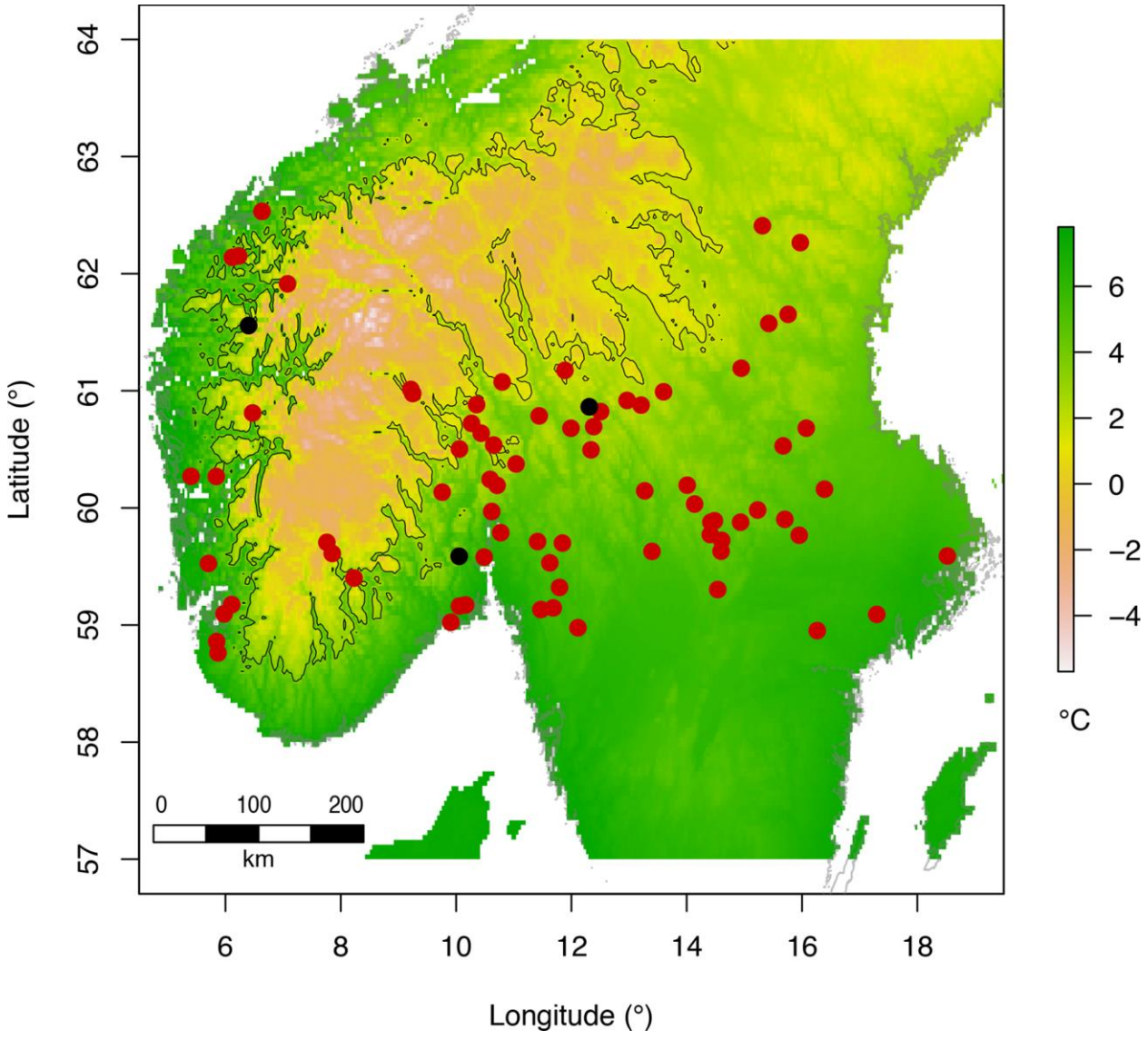
827 **Fig. 4.** Rarefaction curves for 74 samples. Species accumulation curve (obtained by randomizing
828 74 samples) for OTUs against sampling effort are presented in the insert. Dashed line represents
829 the rarefaction sub-sampling level (1000 reads per sample).

830

831 **Fig. 5.** Non-metric multidimensional scaling ordination (NMDS; stress = 0.195), based on Bray-
832 Curtis dissimilarities between protistan communities. Each point represents a lake, with point
833 size scaled by the rarefied OTU richness of the site. A) Arrows represent fitted gradient vectors

834 for spatial (longitude, latitude and altitude) and environmental (log transformed TOC, TP and
835 conductivity) variables. B) Arrows represent corresponding gradient vectors for total and group-
836 specific OTU richness, based on GLMs of the quasi-poisson family. Taxonomic group names are
837 only shown for vectors with permutation-based, FDR-adjusted significance probabilities < 0.05.
838

839 **Fig. 1.** A map of sampled lakes (n=77) colored by annual mean air temperature (ºC) based on the
840 WorldClim database (Hijmans et al., 2005). Contour lines represent altitude below 600 m. The
841 mountain ridge extends S-N around 8ºE. Three lakes indicated by black color were not included
842 in the analysis due to glacial influence or high/low pH.
843



844
845

Fig. 2. Distribution of OTUs and reads (log-scale) of all detected groups and lineages of unresolved phylogenetic placement. Each point represents a group/lineage, and point color reflects taxonomic affiliation (supergroup): yellow = Archaeplastida, red = Alveolata, green = Opisthokonta, blue = Stramenopiles.

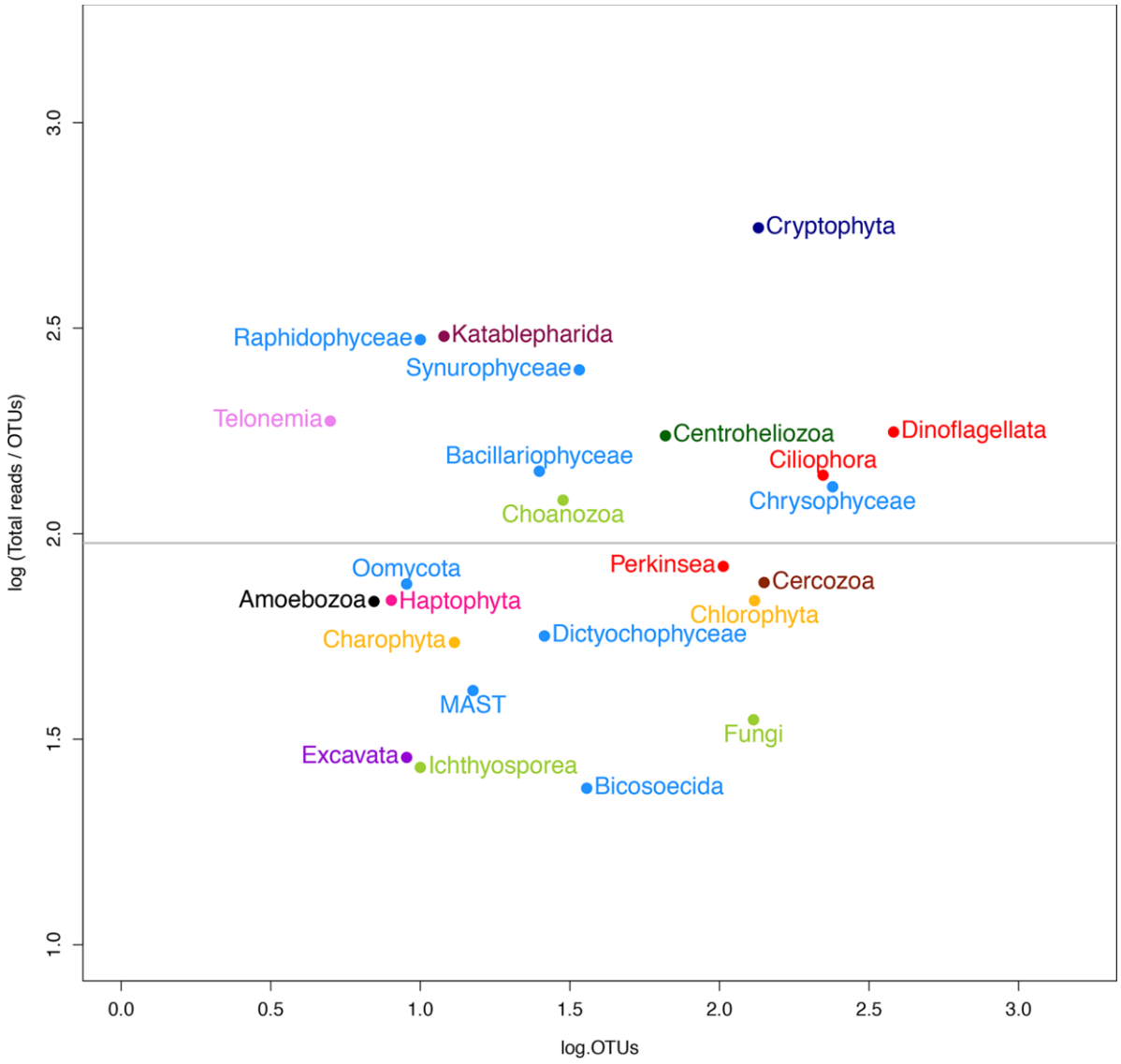
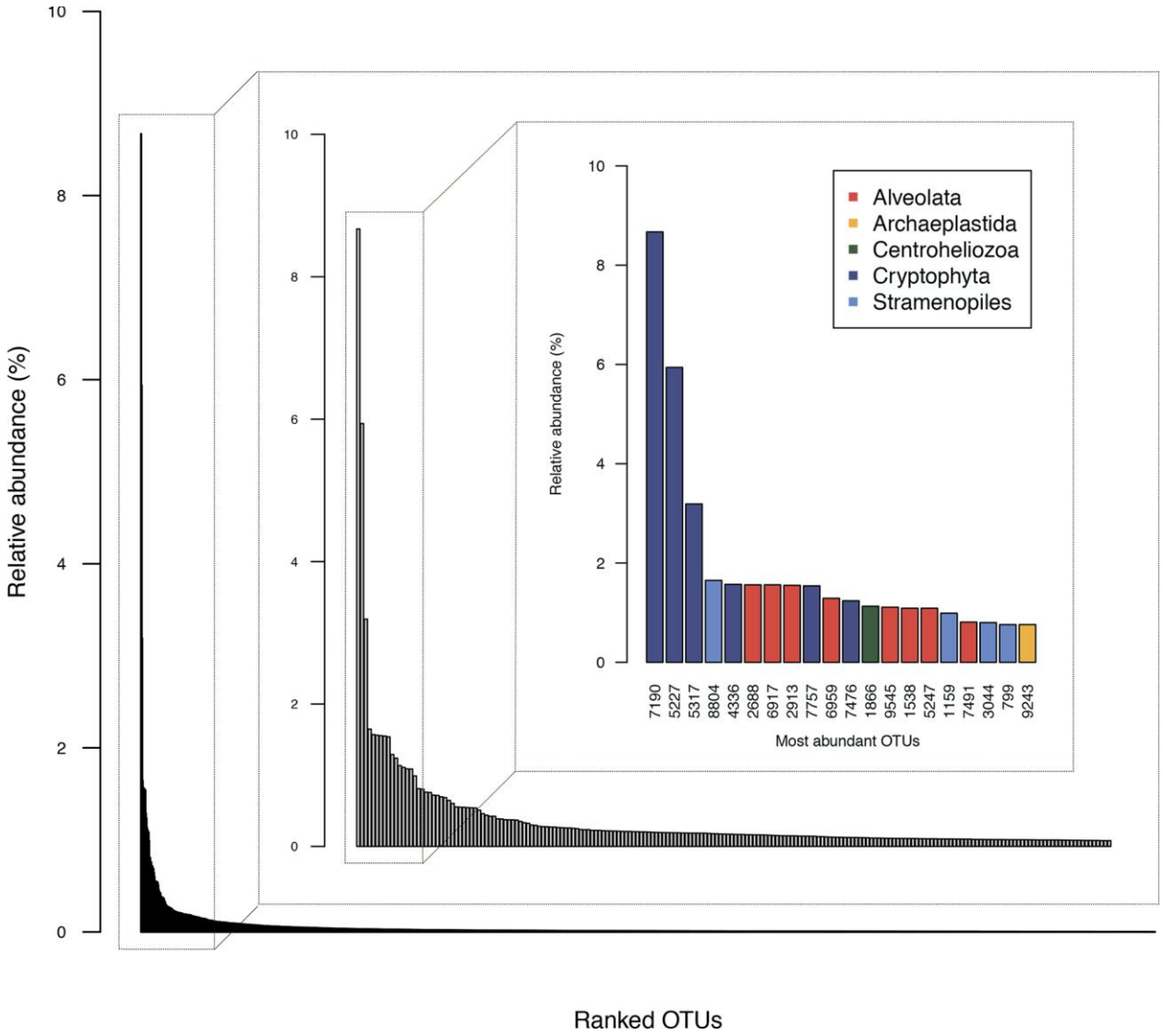
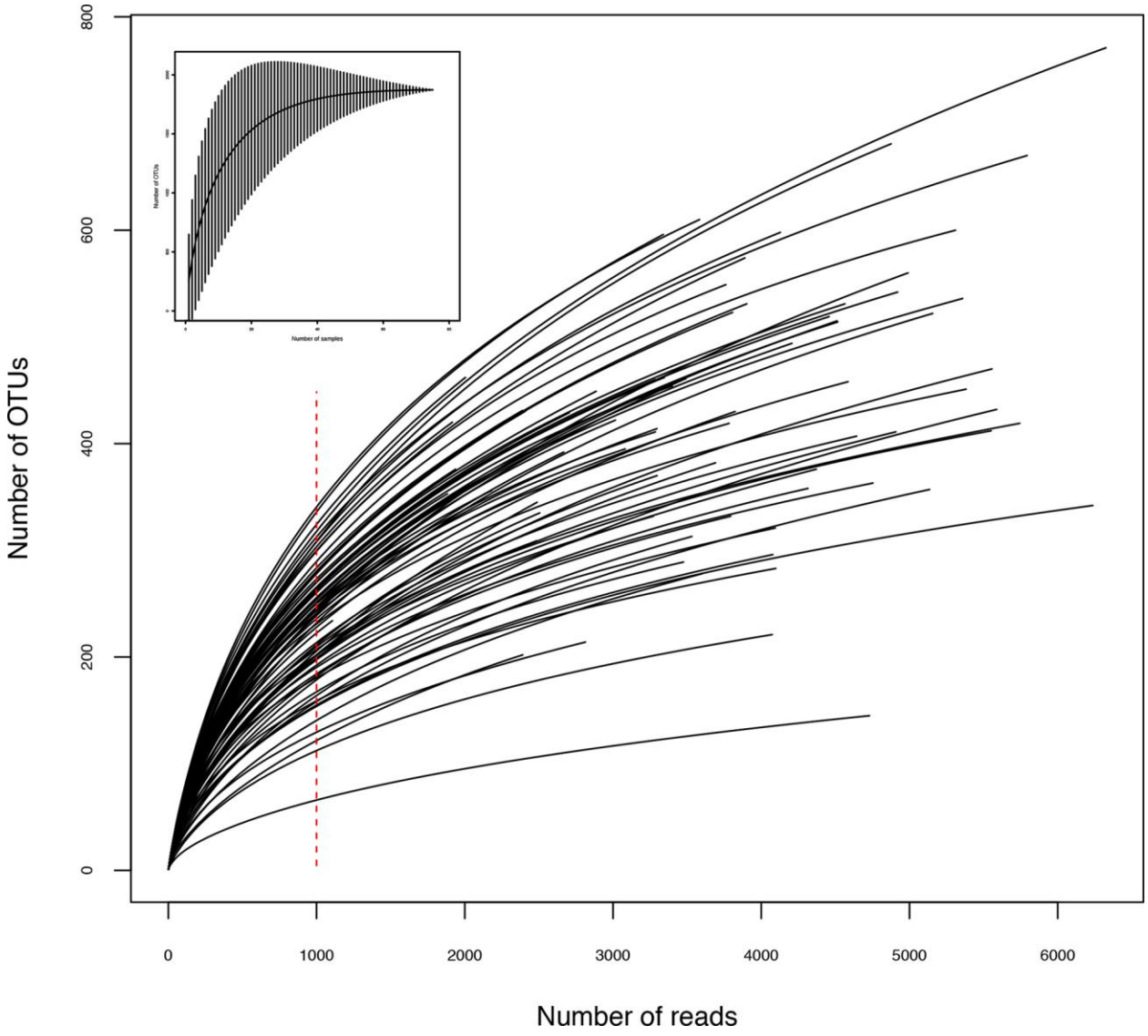


Fig. 3. Rank-abundance curve for the total 1882 OTUs detected in 74 lakes. The relative abundance of top 20 protistan OTUs is shown in the insert. The identity number of the respective OTU is shown below the bars. Colors represent the supergroups: yellow = Archaeplastida, red = Alveolata, green = Opisthokonta, blue = Stramenopiles.



860 **Fig. 4.** Rarefaction curves for 74 samples. Species accumulation curve (obtained by randomizing
861 74 samples) for OTUs against sampling effort are presented in the insert. Dashed line represents
862 the rarefaction sub-sampling level (1000 reads per sample).
863



864

Fig. 5. Non-metric multidimensional scaling ordination (NMDS; stress = 0.195), based on Bray-Curtis dissimilarities between protistan communities. Each point represents a lake, with point size scaled by the rarefied OTU richness of the site. **A)** Arrows represent fitted gradient vectors for spatial (longitude, latitude and altitude) and environmental (log transformed TOC, TP and conductivity) variables. **B)** Arrows represent corresponding gradient vectors for total and group-specific OTU richness, based on GLMs of the quasi-poisson family. Taxonomic group names are only shown for vectors with permutation-based, FDR-adjusted significance probabilities < 0.05.