Title

Protist diversity and metabolic strategy in freshwater lakes are shaped by trophic state and watershed land use at a continental scale

Running title

Continental distributions of freshwater lake protists

Authors and affiliations

Rebecca E. Garner[a,f]#, Susanne A. Kraemer[a,b,c,f], Vera E. Onana[a,f], Yannick Huot[d,f], Irene Gregory-Eaves[e,f], David A. Walsh[a,f]#

[a]Department of Biology, Concordia University, Montreal, Quebec, Canada

[b]Genome Center, Department of Microbiology & Immunology, McGill University, Montreal, Quebec, Canada

[c]Department of Civil Engineering, McGill University, Montreal, Quebec, Canada

[d]Département de géomatique appliquée, Université de Sherbrooke, Sherbrooke, Quebec, Canada

[e]Department of Biology, McGill University, Montreal, Quebec, Canada

[f]Groupe de recherche interuniversitaire en limnologie, Quebec, Canada

#Address correspondence to Rebecca E. Garner, rebecca.garner@mail.concordia.ca or David A. Walsh, david.walsh@concordia.ca.

Word counts

Abstract: 239

Importance statement: 137

Main text: 4,579

Abstract

Protists play key roles in aquatic food webs as primary producers, predators, nutrient recyclers, and symbionts. Yet, a comprehensive view of protist diversity in freshwaters has been challenged by the immense environmental heterogeneity among lakes worldwide. We assessed protist diversity in the surface waters of 366 freshwater lakes across a north temperate to subarctic extent covering nearly 8.4 million km^2 of Canada. Sampled lakes represented broad gradients in size, trophic state, and watershed land use. Hypereutrophic lakes contained the least diverse and most distinct protist communities relative to nutrient-poor lakes. Greater taxonomic variation among eutrophic lakes was mainly a product of heterotroph and mixotroph diversity, whereas phototroph assemblages were more similar under high-nutrient conditions. Overall, local physicochemical factors, particularly ion and nutrient concentrations, elicited the strongest responses in community structure, far outweighing the effects of geographic gradients. Despite their contrasting distribution patterns, obligate phototroph and heterotroph turnover was predicted by an overlapping set of environmental factors, while the metabolic plasticity of mixotrophs may have made them less predictable. Notably, protist diversity was associated with variation in watershed soil pH and agricultural crop coverage, pointing to human impact on the land-water interface that has not been previously identified in studies at smaller scales. Our study exposes the importance of both within-lake and external watershed characteristics in explaining protist diversity and biogeography, critical information in further developing an understanding of how freshwater lakes and their watersheds are impacted by anthropogenic stressors.

Importance

Freshwater lakes are experiencing rapid changes under accelerated anthropogenic stress and a warming climate. Microorganisms underpin aquatic food webs, yet little is known about how freshwater microbial communities are responding to human impact. Here, we assessed the diversity of protists and their myriad ecological roles in lakes varying in size across watersheds experiencing a range of land use pressures by leveraging data from a continental-scale survey of Canadian lakes. We found evidence of human impact on protist assemblages through an association with lake trophic state and extending to agricultural activity and soil characteristics in the surrounding watershed. Furthermore, trophic state

appeared to explain the distributions of phototrophic and heterotrophic protists in contrasting ways. Our findings highlight the vulnerability of lake ecosystems to increased land use and the importance of assessing terrestrial interfaces to elucidate freshwater ecosystem dynamics.

Keywords

microbial eukaryotes; trophic state; plankton; phototrophy; heterotrophy; mixotrophy; human impact

Introduction

Protists have evolved a vast morphological and ecological diversity (1). In aquatic ecosystems, protists play key roles in the transfer of energy and nutrients by converting sunlight into chemical energy, remineralizing organic matter, controlling microbial biomass, feeding higher trophic levels, and maintaining symbioses, some of which recruit prokaryotic metabolisms (2–4).

Elucidating protist diversity in lakes is relevant for clarifying microbial distributions across a wide array of environmental conditions and for tracking the health of critical freshwaters. Covering <1% of Earth's surface (5), lakes contribute disproportionately to the global carbon cycle (6–8) and hold essential water resources (9). Lakes display a rich environmental heterogeneity, generated by integrating fluxes of materials and energy from their catchments and airsheds (10, 11). Freshwater lakes are hotspots of biodiversity, collectively containing higher levels of eukaryote richness and endemism (12) and protist community turnover (13) than the marine and terrestrial realms. In the Anthropocene, lakes are increasingly altered by eutrophication (14), warming temperatures (15), deoxygenation (16), salinization (17), and myriad other persistent and emergent stressors (18). There is accumulating evidence that anthropogenic modifications to lake habitats affect protist assemblages (19–21), in turn influencing ecosystem dynamics.

Efforts to study protists from a variety of biomes have unearthed a vast diversity and begun to map their distributions at a global scale (22–24). Such investigations have brought insights into protist environmental preferences and community assembly processes (25–28), food web dynamics (29–33), symbioses (34–36), viruses (37), functional traits (38), and bioindicator values (39, 40). Large-scale lake surveys have shown that protist assemblages are shaped by broad biogeographic patterns (41–43) but

are also influenced by local environmental factors and interactions with bacteria (44). Phytoplankton surveys recapitulate these observations, while exclusively investigating photosynthetic taxa including Cyanobacteria (45–47). However, there is no clear understanding yet of the distributions of different trophic life history strategies and the environmental drivers underlying their diversity. This knowledge gap is particularly glaring since heterotrophy is likely the most abundant trophic mode (48). Meanwhile, mixotrophs have dramatically altered our view of plankton food webs by combining primary production and prey consumption (49), sometimes surpassing obligate heterotrophs as the leading grazers of bacteria in lakes (50, 51).

In this study, 18S rRNA gene amplicon sequencing was used to investigate the distributions of protists in the surface waters of 366 freshwater lakes across a north temperate to subarctic continental extent. We hypothesized that (1) protist diversity at the local scale and community turnover decrease under high-nutrient conditions, (2) trophic groups respond to different environmental factors, specifically, phototrophs are more sensitive than heterotrophs to bottom-up resource availability, and (3) because lakes integrate their catchments, protist diversity in lakes should reflect watershed conditions. This project was conducted within the LakePulse survey, which sampled hundreds of lakes of different sizes in watersheds under varying levels of human impact with the primary aim to assess lake health through a multidisciplinary lens (52). The current study fills a gap in the mapping of microbial biogeography through this first standardized assessment of protist diversity across Canada, which stewards the greatest abundance of lakes worldwide (53). Our study draws attention to the diversity of protists and the ecological patterns that emerge in a broad collection of newly explored habitats being reshaped by increasing human impact.

Results

Sampled lakes and watersheds display high environmental heterogeneity

Protist assemblages were surveyed in the euphotic zones of 366 freshwater to oligosaline lakes in 12 ecozones across Canada (43 – 68 °N, 53 – 141 °W) (Figure 1). Watersheds ranged widely in area (0.3 – 9,332.3 km^2) and were characterized by a variety of human population densities (0 – 3,785 people/km^2) and land use, including different proportions of crop agriculture (0 – 81%) and built development (0 –

93%). Lakes had a wide range of surface areas (0.05 – 99.66 km^2) and maximum depths (1 – >150 m) and were either vertically mixed (~40% of lakes) or thermally stratified (~60%) at the time of sampling. Physicochemical conditions differed substantially across lakes, as represented by a broad pH gradient (5.6 – 10.2) and ultraoligotrophic to hypereutrophic states evaluated by total phosphorus (TP) concentrations (2 – 2,484 µg/L) (Figure 1).

Sampled lakes and watersheds reflected regional variation in environmental conditions including anthropogenic gradients (Figure S1). Lakes in northern Canada (Taiga Cordillera, Boreal Cordillera, and Taiga Plains ecozones) were subject to the coldest climates and lowest intensities and proportions of land use within their watersheds. Lakes in western Canada (Montane Cordillera, Pacific Maritime, and Semi-Arid Plateaux) were the deepest on average and located in watersheds with the largest proportions of harvested forests. In central Canada (Boreal Plains and Prairies), a region dominated by plains and prairies with intensive agriculture, lakes were generally shallow, exposed to winds, productive, and high in pH, carbon, ions, and nutrients. Lakes in eastern Canada (Mixedwood Plains, Boreal Shield, Atlantic Highlands, and Atlantic Maritime) had the warmest surface waters and generally the most built-up landscapes, a feature shared with watersheds in the Pacific Maritime. Overall, hypereutrophic, ion- and carbon-rich, and high-pH lake conditions were most often observed in agricultural watersheds with alkaline soils (Figure S2).

Lakes support taxonomically and functionally diverse protist assemblages

Eukaryotic diversity was assessed through the sequencing of 18S rRNA gene fragments amplified from DNA collected in 0.22 – 100 µm surface water particles. A total of 15,848 amplicon sequence variants (ASVs) were inferred in 17,749,930 sequences across 366 lake samples. A final data set of 13,046 putative protist ASVs encompassing 14,622,273 sequences was retained after ASVs assigned to animals, fungi, and plants were removed (Table S1). The rarefaction of pooled samples showed that the sampling of new ASVs plateaued toward 2,000,000 sequences, signaling that the global sequencing effort had exhaustively captured the protist diversity targeted by the primer pair (Figure S3). The most abundant taxonomic groups were Ochrophyta (24% of all sequences), Cryptophyta (18%), Ciliophora

(15%), and Dinoflagellata (11%) (Figure 1). Lineages with the highest ASV richness were Ochrophyta (22% of all ASVs), Dinoflagellata (12%), and Chlorophyta (11%).

We analyzed ASV incidence to assess the contribution of individual assemblages to total landscape diversity. New genotypes accumulated at a high rate in the first ~100 randomly ordered assemblages, followed by a gradual deceleration (Figure 2A). The majority of ASVs were restricted to one or a few lakes (Figure 2B). A smaller number of ASVs were distributed widely, including one ASV assigned to *Cryptomonas curvata* (Cryptophyta) that was ubiquitous yet highly variable in relative abundance (0.0031 – 64%) across all the lakes sampled.

Trophic functions were assigned to ASVs representing 85% of the total sequence space by leveraging natural history descriptions summarized from the literature (Figure 1; Table S1). Of the functionally annotated ASVs, most were classified as obligate phototrophs (32% of all sequences). The most abundant phototrophs were classified under Chrysophyceae (Ochrophyta; 16% of all sequences), Bacillariophyta (Ochrophyta; 3%), and Chlorophyceae (Chlorophyta; 5%). Bacterivory (20% of all sequences) and mixotrophy (20%) were the next most abundant trophic modes. The most abundant mixotrophs were classified under Cryptophyceae (Cryptophyta; 17% of all sequences) and Dinophyceae (Dinoflagellata; 2%). Cytotrophy (i.e. feeding on other protists), parasitism, commensalism, saprotrophy, and osmotrophy were detected to a lesser extent. Parasites (3% of all sequences) were most abundant in the Coccidiomorphea (Apicomplexa; 2%) and Oomycota (Pseudofungi; 1%). Heterotrophs – which broadly encompassed bacterivores, cytotrophs, saprotrophs, and osmotrophs – comprised 27% of all sequences. Heterotrophy was most abundant in the ciliates Spirotrichea (9% of all sequences), Oligohymenophorea (2%), and Litostomatea (2%) and in other lineages including Bicoecea (Opalozoa; 3%) and Katablepharidaceae (Katablepharidophyta; 2%).

Genotypic similarity to known protist 18S rRNA gene diversity was assessed through the global alignment of ASVs to V7 region fragments in the Protist Ribosomal Reference database (PR^2) (Figure S4). Of the 12,511 ASVs that met the threshold global alignment length, the majority showed high sequence similarity with references in the database. Most ASVs (58%) were ≥96% identical to references. ASVs with 100% similarity to reference sequences occupied 56% of the total sequence space, while ASVs with ≥96% similarity occupied 91% of sequences. The most novel genotypes included the 738

ASVs with <90% sequence similarity to PR^2 references and were either not assigned to a supergroup (415 ASVs) or primarily classified as Opisthokonta (141), Alveolata (88), or Stramenopiles (30).

Local diversity at each lake was estimated by richness (67 – 1,275 ASVs) (Figure S5), the Shannon index (0.17 – 5.67; Figure 3A), Pielou's evenness index (0.04 – 0.83), and Faith's phylogenetic diversity index (8.15 – 76.04). The Shannon index was negatively correlated with magnesium (r = -0.33), total nitrogen (TN; r = -0.32), dissolved inorganic carbon (DIC; r = -0.31), and potassium (r = -0.30) concentrations (all correlations had $p < 0.001$). Tukey's tests of ANOVAs comparing the association of trophic state with local diversity showed that mean richness, Shannon diversity, and evenness were significantly lower in hypereutrophic lakes than in eutrophic, mesoeutrophic, mesotrophic, or oligotrophic lakes (all $p \leq 0.002$) (Figure 3B).

Protist assemblages vary regionally and across lake trophic states

Next, we looked at how communities varied in taxonomic and phylogenetic composition among lakes. A principal component analysis (PCA) of ASV assemblages showed a clear pattern of taxonomic variation by lake trophic state and ecozone along the first dimension (Figure 4A). Assemblages in the typically nutrient- and ion-rich lakes of the Prairies and Boreal Plains were distinguished from assemblages in the lower-nutrient lakes of the Boreal Shield and other eastern regions. Cryptophyte diversity (*Cryptomonas*, *Geminigera*, *Plagioselmis*, and *Komma* species) contributed the most strongly to the variation among assemblages (Figure 4A). A principal coordinate analysis (PCoA) of generalized UniFrac distances between assemblages showed patterns of phylogenetic variation that were highly congruent with the taxonomic variation observed in the PCA, as evaluated by an RV coefficient correlating the first two dimensions of each ordination (RV = 67%, $p = 1.3 \times 10^{-98}$) (Figure 4B).

Taxonomically distinct assemblages as quantified by local contributions to β-diversity (LCBD) were mostly localized in the Prairies and Boreal Plains, with a few high-LCBD assemblages scattered across other regions (Figure 5). β-diversity partitioning showed that the taxonomic dissimilarities between assemblages were primarily generated through ASV turnover (75% of total variance) and differences in richness to a lesser extent (25%). Highly significant ($p < 0.001$) positive correlations were detected between LCBD and a complement of physicochemical variables including potassium (r = 0.53), TN (r =

196 0.50), DIC (r = 0.45), TP (r = 0.44), dissolved organic carbon (DOC; r = 0.42), magnesium (r = 0.40),
197 sodium (r = 0.38), sulfate (r = 0.37), chlorophyll-*a* (r = 0.33), lake colour (r = 0.22), and pH (r = 0.22).
198 LCBD further correlated (p < 0.001) with watershed crop cover (r = 0.35) and soil pH (r = 0.27) and
199 correlated negatively with soil nitrogen (r = -0.25). LCBD was negatively correlated with local diversity,
200 estimated as ASV richness (r = -0.44) and the Shannon (r = -0.54) and Faith's phylogenetic diversity (r = -
201 0.34) indices (all p < 0.001).
202
203 *Physicochemical and watershed conditions predict community turnover*
204 Following the previous observations of high community variability across an environmentally
205 heterogeneous set of lakes, we evaluated the drivers of taxonomic and phylogenetic turnover based on
206 five categories of environmental variables: (1) geography (i.e. latitude, longitude, and altitude), (2)
207 weather, (3) lake morphometry, (4) physicochemistry, and (5) watershed characteristics including land
208 use and edaphic properties.
209 We employed generalized dissimilarity models (GDMs) to detect nonlinear trends between
210 community turnover and environmental gradients. Physicochemical factors explained the highest
211 deviance in GDMs modeling taxonomic or phylogenetic turnover. In order of decreasing strength, DIC,
212 TP, chlorophyll-*a*, magnesium, pH, potassium, lake colour, surface temperature, and DOC were
213 statistically significant predictors of taxonomic turnover, whereas chlorophyll-*a*, magnesium, TN, pH,
214 calcium, and colour were significant predictors of phylogenetic turnover. After physicochemistry,
215 watershed characteristics were the most important predictors of community turnover, with both taxonomic
216 and phylogenetic diversity responding most strongly to soil pH and then proportion of cropland cover. The
217 volumetric fraction of soil coarse fragments and natural landscape coverage were additional predictors of
218 taxonomic and phylogenetic turnover, respectively. Lake morphometry, comprising maximum depth,
219 watershed slope, and shoreline circularity, had relatively weak effects on turnover. Weather and
220 geography did not generate statistically significant GDMs. Model deviances are summarized in Table 1
221 and the partial effects of individual variables are summarized in Figure S6 and Table S2.
222
223 *Trophic strategies exhibit contrasting distributions*

To examine the distributions of different trophic strategies, we investigated the taxonomic variation of phototroph, heterotroph, and mixotroph communities in separate PCAs (Figure 6). Reflecting the taxonomic variation emerging at the whole-community level, assemblages of each trophic mode were distinguished by lake trophic state and ecozone along the first dimension. However, phototrophs displayed contrasting taxonomic variation patterns to heterotrophs and mixotrophs. To compare the taxonomic turnover of phototrophs, heterotrophs, or mixotrophs among lakes of the same trophic state, we measured the distances of assemblages to the trophic state median within the two-dimensional principal coordinate space. The mean distance of phototroph assemblages to the trophic state median (i.e. turnover) was significantly lower in hypereutrophic lakes than in eutrophic, mesoeutrophic, mesotrophic, or oligotrophic lakes (all $p < 0.001$) (Figure S7). In contrast, the mean distance of heterotroph assemblages was significantly higher in eutrophic lakes than in mesotrophic or oligotrophic lakes (all $p \leq 0.007$) (Figure S7). Compared with either phototroph or heterotroph assemblages, mixotroph assemblages were highly dispersed within trophic state groups (Figure S7).

Partitioning protist assemblages by trophic function further allowed us to examine the responses of different groups to environmental conditions. GDMs showed that phototrophs, heterotrophs, and mixotrophs each responded most strongly to physicochemical gradients. Chlorophyll-*a*, DIC, potassium, and pH were predictors common to the three trophic modes, while calcium, colour, and chloride were additional predictors of phototroph turnover, and TP, surface temperature, sulfate, TN, colour, and chloride were additional predictors of heterotroph turnover. Chlorophyll-*a* followed by potassium were the top predictors of phototroph turnover, while DIC followed by chlorophyll-*a* were the top predictors of heterotroph and mixotroph turnover. Watershed characteristics explained an important amount of deviance in phototroph and heterotroph turnover: phototroph turnover was explained by soil pH, crop coverage, soil organic carbon density, and built development, whereas heterotroph turnover was additionally explained by soil coarse fragments. Only phototroph turnover was predicted by lake morphometry variables (watershed slope, lake maximum depth, and circularity). Heterotroph turnover was weakly explained by weather, specifically air temperature and wind speed, and geography, specifically altitude and distances between lakes. Mixotroph turnover was the least predictable from environmental conditions, as physicochemical factors explained less deviance than for obligate phototroph or

252 heterotroph turnover and GDMs modeled on geography, weather, lake morphometry, and watershed
253 characteristics were not statistically significant. Model deviances are summarized in Table 1 and the
254 partial effects of individual variables are summarized in Figure S8 and Table S2.
255
256 **Discussion**
257 *Protist diversity unveiled at a continental scale*
258 Establishing a comprehensive perspective of freshwater protist diversity is challenging given the
259 substantial environmental heterogeneity of the millions of lakes distributed globally. Our study fills a
260 sizable gap by mapping protist distributions across hundreds of lakes spanning an area covering nearly
261 8.4 million km^2 in the largest study of its kind to employ a standardized sampling scheme (52). Protist
262 diversity was determined in the sunlit surface waters of 366 freshwater lakes varying in size and degree
263 of human impact on the watershed. The sampled biomass bridged cell diameters across four orders of
264 magnitude (0.22 – 100 µm), allowing us to recover pico- to microscale organisms, many of which are not
265 resolved under standard light microscopy and which constitute an expansive microbial diversity
266 encompassing the major eukaryotic lineages and trophic strategies. Our global rarefaction analysis
267 showed that the sequencing effort provided a reasonable estimate of the genotypic variation captured by
268 the primer pair. The rapid accumulation of genotypes across sites indicated that sampling hundreds of
269 lakes was required to assess landscape diversity. The integration of our work with the collection of recent
270 large-scale surveys is leading to a synoptic view of protist ecology (23, 24, 28, 42, 43, 54–57).
271
272 *Surface water assemblages contain high proportions of phototrophs and heterotrophs*
273 Lacustrine protist diversity was dominated by ochrophytes, which accounted for both the highest
274 sequence abundance and ASV richness. Cryptophytes, ciliates, dinoflagellates, and chlorophytes were
275 also highly represented, reflecting taxonomic profiles typical of freshwater biomes on other continents and
276 identified with primer pairs targeting other gene regions (58). The main taxa contributing to the
277 dissimilarity between assemblages were mixotrophic cryptophytes, whose distributions as major
278 bacterivores may be dependent on the occurrence of specific prey (59).

Phototrophic taxa accounted for the greatest richness and relative abundance of ASVs, perhaps not unexpectedly given that assemblages were sampled from the euphotic zone. Notably, the dominance of phototrophs within the protist fraction appears to be a distinct feature of freshwater photic zones in contrast with the prevalence of heterotrophs in the sunlit ocean and surface soils (13). We showed that heterotrophs and mixotrophs were numerically important groups after phototrophs, which, along with a smaller collection of parasites, illustrates that lake surface waters harbour a broad array of microbial functions linking multiple trophic levels.

Protist communities respond to local environmental conditions

The key environmental drivers of protist diversity in Canadian lakes departed from those observed in previous large-scale surveys. Assemblages showed the strongest responses to physicochemical factors, including nutrient, major ion, and chlorophyll-*a* concentrations, pH, and lake colour. The environmental drivers in lakes differed from those in marine and soil ecosystems, where protist diversity at large scale is generally predicted by temperature (60) and annual precipitation (28), respectively. The differences in environmental filtering between biomes likely represent fundamental differences in the types of habitats and degrees of ecosystem connectivity. Compared with the more spatially continuous and expansive ocean and soil macroenvironments, lakes are fragmented across the landscape and neighbouring lakes can exhibit widely contrasting physicochemical attributes (10, 61). Lake heterogeneity is amplified by temporal variability and punctuated perturbations. Higher community turnover has previously been measured among lakes than within marine or soil ecosystems (12, 13), which we reassert is linked to physicochemical heterogeneity.

We found that the influence of local environmental conditions far outweighed the effects of geographic variation across the continental extent. Among LakePulse sites, systems with the highest trophic states (typically, Prairies and Boreal Plains lakes) were located at intermediate longitudes, latitudes, and altitudes. A study of Scandinavian boreal lakes spanning a longitudinally-aligned and narrower trophic state gradient (oligotrophic to mesoeutrophic) reported that geography explained more protist community variation than water chemistry (42), complementing our assessment that regional physicochemical heterogeneity is a major determinant of protist diversity. Other surveys identified

biogeographic patterns of protist diversity structured by the isolation and dispersal limitation of mountain lake communities (41, 43). Geographic barriers (e.g., the Rocky Mountains) did not appear to generate strong compositional divisions in our set of protist assemblages but were identified as having an important influence on the distributions across LakePulse sites of crustacean zooplankton (62), a group with greater dispersal limitation due to their larger body sizes. Instead, we found the greatest taxonomic and phylogenetic divisions between regions distinguished by differences in lake trophic state and other local environmental conditions.

Partitioning protist trophic diversity allowed us to examine how different components of freshwater food webs respond to the environment. We observed contrasting distribution patterns for each trophic mode. Phototroph assemblages among hypereutrophic lakes exhibited significantly lower taxonomic turnover (as evaluated by mean distance to the trophic state median) than lakes at lower nutrient states, whereas heterotroph assemblages did not follow this trend but turned over significantly more rapidly among eutrophic lakes than mesotrophic or oligotrophic lakes. Yet, the turnover within each trophic mode was predicted by a mostly overlapping suite of physicochemical factors, including chlorophyll-*a*, DIC, ions, pH, and colour, although rank order of importance varied among groups. Nutrient (TP and TN) concentrations and surface temperature were exclusive predictors of heterotroph turnover. All of the environmental predictors of mixotroph turnover were common to both phototrophs and heterotrophs, with no predictors unique to mixotrophs. Changes in low levels of chlorophyll-*a* were associated with the most rapid turnover in phototroph composition, which is to be expected given that chlorophyll-*a* is linked to phytoplankton biomass and phototroph diversity shifts along a lake productivity gradient. The next strongest predictor of phototroph turnover was potassium, which is not a limiting resource (63) but displayed extreme regional variation peaking in the Prairies, followed by the Boreal Plains. Heterotroph and mixotroph turnover was primarily predicted by DIC concentrations. Overall, the environmental drivers of heterotroph diversity mostly reflect the bottom-up controls on primary producers traversing multiple trophic levels but are rendered further complex by the added effects of nutrient and temperature factors. Bacterial prey and top-down controls, encompassing predation and parasitism (not measured in this study), likely also determine trophic functional diversity.

Mixotroph distributions were the least aligned with trophic state and the least predictable from environmental conditions, which should be expected for organisms with the metabolic versatility to occupy variable niche spaces. The balance between primary production and prey consumption is dependent on a mixotroph's phenotypic plasticity (64). While mixotrophy is competitively advantageous over obligate phototrophy or heterotrophy under low-nutrient conditions (65), primarily phototrophic mixotrophs prevail in oligotrophic lakes and are replaced by primarily heterotrophic mixotrophs as trophic state increases (66). Another variable driving mixotroph diversity in aquatic ecosystems is light availability (67, 68), which is modulated by lake colour, a predictor identified in this study. Here too, bacterial prey and zooplankton predators with a preference for nutritious mixotrophs likely also exert controls (69).

Hypereutrophic lakes are taxonomically distinct

Hypereutrophic lakes, located mostly in agricultural watersheds, contained protist assemblages with the lowest diversity and highest taxonomic distinctness (i.e. LCBD) relative to other lakes in the landscape. Specifically, Shannon diversity was inversely related to ion and nutrient concentrations, while taxonomic distinctness tracked with ion- and nutrient-rich conditions. Hypereutrophic conditions potentially filter protist communities by creating relatively extreme conditions (e.g., light attenuation), tolerated by a small number of taxa assembling into uneven communities distinct from those in lakes at lower nutrient states. Phosphorus is often the limiting nutrient of phytoplankton in numerous freshwater systems (70, 71), yet TP was a predictor of turnover exclusive to heterotrophs. Phosphorus was found to be an important predictor of littoral protist diversity in European lakes (44) and of long-term microeukaryote community turnover evidenced from paleolimnological trends (19). Interactions between protists and Cyanobacteria likely also play a role in determining protist assemblages in hypereutrophic lakes, especially as Cyanobacteria of the order *Microcystis* were found to be associated with high-nutrient conditions across LakePulse sites (72).

Ordinations of separate trophic modes showed that the high community variation among eutrophic lakes was generated by heterotroph diversity, whereas obligate phototroph assemblages were the least varied under hypereutrophic conditions. A positive relationship between compositional heterogeneity and trophic state runs counter to the expectation of reduced community variability (e.g., for

362 phytoplankton (73)) that is predicted to follow the leveling of abiotic conditions among lakes induced by
363 land use and eutrophication. Biotic homogenization appears to have trended with long-term climate
364 warming and eutrophication in Cyanobacteria (74) and protist (20) assemblages reconstructed from
365 sediment core chronologies. Following our observation of increased compositional heterogeneity among
366 lakes as a function of trophic state, we posit that protist communities in productive lakes are less stable
367 over time, including over the same season as observed in bacterial time series (75). The temporal fluxes
368 in abiotic conditions prompting succession may be induced by allochthonous inputs or nutrient
369 resuspension from sediments accrued at higher rates in regions of extensive lake use and high
370 populations densities (76). Furthermore, given that many Prairies and Boreal Plains lakes are shallow and
371 exposed to winds, temporary stratification followed by destratification is not uncommon (77). We
372 speculate that taxon replacement linked to land use and eutrophication may force a re-evaluation of
373 human impact on biodiversity. In particular, anthropogenic pressures may not inherently decrease
374 diversity but instead increase turnover, possibly at the expense of rare or specialist taxa disappearing
375 from the landscape pool (78).
376
377 *Watersheds influence lacustrine protist diversity*
378 While the importance of physicochemical factors on lacustrine protist diversity has been described (42,
379 44) and elaborated upon in this study, the influence of the watershed, in particular soil properties and land
380 use that are often but not entirely correlated with lake physicochemical attributes, until now have not been
381 documented at the continental scale. We found that the taxonomic distinctness of local assemblages (i.e.
382 LCBD) corresponded strongly with the proportion of crop agriculture in the watershed, while surface soil
383 pH was an important predictor of community turnover. Because of their concave topographies and
384 position in the landscape, lakes are recipients of major allochthonous subsidies, with global effects (e.g.,
385 carbon storage) that are disproportionate to the spatial extent of lakes (6). The influence of terrestrial
386 catchments on within-lake community dynamics is compelling evidence for why lakes cannot be studied
387 in isolation of their watersheds.
388 The filtering of lacustrine protist diversity by watershed soil chemistry points to the influence of
389 external factors on lake conditions. Soil buffering capacity, determined by soil texture, organic matter

content, and mineral composition, is a main abiotic control on lake water pH (79). Furthermore, soil properties control the mobility of nutrients and their eventual input into lakes. Specifically, soil pH determines the availability and chemical forms of nutrients, and particle size governs the movement of groundwater carrying released nutrients (80). Soil properties also determine the composition and activity of soil microbiomes, which perform biogeochemical transformations modulating the availability of nutrients (81) and seed potential colonists from soils to lakes. Altered precipitation regimes and warming temperatures associated with climate change are expected to increase soil erosion (82) as well as terrestrial nutrient exports (83).

Given the continual increase in land surface transformed by agricultural production and urbanization (84), accelerated watershed land use conversions are widespread. Changes in soils associated with human activities range from increased nutrient loading and acidification by nitrogenous fertilizers in agriculture (85) to shifts in carbon storage precipitated by changes in land management practices or climate (86). Moreover, land use and climate change interact to increase the frequency and magnitude of nutrient and carbon pulses to waterbodies (87). In the interest of securing a healthy future for critical freshwaters, we suggest that current soil chemistry heterogeneity across the landscape can inform predictions about the potential consequences of anthropogenic watershed alterations on lakes. In particular, work can be done to understand the microbial diversity and food web dynamics driven by various soil states under future land use scenarios. Overall, the ability to predict lacustrine protist diversity from watershed conditions, as demonstrated in this study, highlights an expanded potential for monitoring lake ecosystems using remote sensing products (88).

Conclusion

This is the first study to examine the taxonomic and trophic functional variation in protist diversity across the expansive and lake-rich Canadian landscape. We showed that lakes at this continental scale displayed broad environmental heterogeneity including substantial variation in local physicochemical conditions driving taxonomic and phylogenetic community turnover. Watershed soil pH and crop agriculture additionally predicted community turnover and exceptional local-scale diversity. Hypereutrophic lakes were found to contain less diverse and more distinct assemblages than lower-

418 nutrient lakes, primarily as a product of their variable heterotroph and mixotroph compositions. In
419 contrast, phototroph assemblages were more similar among hypereutrophic lakes. While phototrophy was
420 the prevailing nutritional strategy in lake euphotic zones, heterotrophy was nearly as numerically
421 important; each of these trophic modes was highly predictable from physicochemical and other
422 environmental factors. Our survey and findings serve as a valuable resource for mapping species
423 distributions and provide a basis for future research into the increasing anthropogenic impact on lake
424 microbiomes.
425
426 **Methods**
427 *Lake selection and sampling*
428 Hundreds of lakes were sampled between July and early September in 2017 – 2019 by the Natural
429 Sciences and Engineering Research Council of Canada (NSERC) Canadian Lake Pulse Network (52).
430 Sampling was timed to coincide with the summertime period of water column thermal stratification, where
431 relevant. Lakes were sampled across 12 terrestrial ecozones, regions defined by landform, geology, and
432 vegetation (89). Lake selection was stratified across lake surface area and watershed land use impact
433 categories to capture natural and human-mediated lake heterogeneity. Only natural lakes with a
434 maximum depth of at least 1 m and within 1 km from a road were considered. Freshwater to oligosaline
435 lakes (identified as having conductivity <8 mS/cm and total major ions <4,000 mg/L) were retained for this
436 analysis.
437 Water was collected using an integrated tube sampler from the euphotic zone over a depth of up
438 to 2 m below the surface at the deepest point in the lake (90). The site of maximum lake depth was
439 located by depth sounding with the aid of bathymetric maps where available. The depth of the euphotic
440 zone was estimated as twice the Secchi disk depth. All water sampling equipment was acid-washed and
441 rinsed three times with lake water before use. Carboys were stored in icepack-chilled coolers until water
442 filtration later in the day. Water was prefiltered through 100 µm nylon mesh and vacuum-filtered on 47
443 mm-diameter 0.22 µm Durapore membranes through a glass funnel at a maximum pressure of 8 inHg. Up
444 to 500 mL of water was filtered until the filter was nearly clogged. Filters were stored in sterile cryovials at
445 -80 °C.

446

447 *18S rRNA gene amplification and sequencing*

448 DNA was extracted using the DNeasy PowerWater kit (QIAGEN, Hilden, Germany) according to the
449 manufacturer's instructions with the addition of two optional steps: after bead beating and centrifugation,
450 1 µL ribonuclease A was added to samples, followed by 30 min incubation at 37 °C. DNA was quantified
451 using the Qubit dsDNA BR Assay (Invitrogen, Carlsbad, CA, USA). A ~265 bp fragment of the 18S rRNA
452 gene V7 region was amplified with the primers 960F (5'-GGCTTAATTTGACTCAACRCG-3') (91) and
453 NSR1438 (5'-GGGCATCACAGACCTGTTAT-3') (92) to broadly target microeukaryotes (93). Each PCR
454 reaction contained a total 25 µL mixture of 14.25 µL MilliQ, 5 µL 5X High-Fidelity buffer, 1.25 µL of each
455 10 µM primer, 0.5 µL 10 mM dNTPs, 0.25 µL DMSO, 0.5 µL Phusion DNA polymerase (Thermo Fisher
456 Scientific, Waltham, MA, USA), and 2 ng DNA template. PCR conditions followed an initial denaturation at
457 98 °C for 1 min, 30 cycles of 98 °C for 10 s, 60 °C melting for 30 s, 72 °C for 20 s, and a final extension at
458 72 °C for 5 min. PCR products were loaded with Orange G dye into ethidium bromide-stained 2%
459 agarose gel and electrophoresed at 40 V for 100 min. DNA bands aligned at the target fragment length
460 against a 100 bp DNA ladder were excised with razor blades and gel-extracted with the QIAquick Gel
461 Extraction kit (QIAGEN, Hilden, Germany), modified by final elution into MilliQ. PCR products were
462 submitted to Genome Quebec for library barcoding and sequencing of 250 bp paired-end reads in three
463 sequencing runs on an Illumina MiSeq platform.

464

465 *ASV inference and annotation*

466 Primer sequences were removed in Cutadapt v. 3.1 (94). Trimmed reads were processed into ASVs
467 through DADA2 v. 1.16 (95). Samples were pooled for ASV inference using otherwise default parameters.
468 Taxonomy was assigned with naïve Bayesian classification trained on PR^2 v. 4.12.0 (96). Potentially
469 spurious ASVs were removed by visually inspecting a *de novo* alignment performed in MAFFT (97). To
470 retain only putative protist ASVs, ASVs assigned to Metazoa, Fungi, and Embryophyceae were removed.
471 Taxa were assigned to trophic functional groups as either photoautotrophs, heterotrophs (bacterivores,
472 cytotrophs, saprotrophs, or osmotrophs), mixotrophs, or parasites according to lineage-specific feeding
473 habits summarized by Adl *et al*. (98).

Sequence similarities with known diversity

A database of 18S rRNA gene references restricted to the V7 region was constructed by applying the 960F/NSR1438 primer pair to PR^2 v. 4.13.0 in Cutadapt. ASV top hits were queried against the PR^2 database in BLAST v. 2.6.0+ (99). Sequence identities were reported for ASVs that were globally aligned to references over a length ≥220 nucleotides.

Environmental data collection

Lake trophic states were assigned based on TP concentration thresholds estimated for Canadian freshwater systems: ultraoligotrophic (TP <4 µg/L), oligotrophic (4 – 10), mesotrophic (10 – 20), mesoeutrophic (20 – 35), eutrophic (35 – 100), and hypereutrophic (>100) (100). Meteorological conditions recorded over seven days leading up to sampling and ice disappearance day data were accessed from ERA5-Land hourly reanalysis (101). Data on watershed slope and lake volume, discharge, and hydraulic residence time were accessed from HydroLAKES v. 1.0 (5). Watershed surface soil properties were accessed from SoilGrids250m (102). Land cover information was compiled as described by Huot *et al*. (52). Maps were constructed in R with the NAD 83 coordinate reference system and using the coordinates of Canada from the package maps (103) and ecozone shapefiles sourced from the Canada Council of Ecological Areas (89).

Environmental data were categorized into thematic groups of variables. Latitude, longitude, and altitude were categorized as geography variables. Ice disappearance day and meteorological variables (air temperature, precipitations, and net solar radiation) were categorized as weather variables. Lake surface area, circularity, volume, maximum depth, discharge, residence time, watershed slope within 100 m of the shoreline, watershed area, and lake-to-watershed area ratio were categorized as lake morphometry variables. Watershed land use (crop agriculture, pasture, built development, and clear-cut forestry) and natural land cover fractions, human population density, and mean surface soil properties (bulk density of the fine earth fraction, cation exchange capacity, nitrogen, pH, organic carbon density, organic carbon content in the fine earth fraction, volumetric fraction of coarse fragments, clay, sand, and silt) were categorized as watershed variables. Surface water temperature, calcium, magnesium,

potassium, sodium, chloride, sulfate, TP, TN, DIC, DOC, and chlorophyll-*a* concentrations, pH, and lake colour were categorized as lake physicochemical variables. Missing physicochemical data were replaced with ecozone median values. Highly collinear variables, evaluated by Pearson's correlation $r \geq 0.7$, within all categories except physicochemistry were removed.

Diversity analyses

ASVs were aligned in the SILVA Incremental Aligner v. 1.7.2 (104) against the SILVA 138.1 SSU Ref NR 99 database (2020/08/27 release) (105). A maximum-likelihood phylogeny was constructed in FastTree v. 2.1.11 using the Generalized Time-Reversible model of nucleotide evolution (106). Phylogenetic dissimilarities between ASV assemblages were calculated as generalized UniFrac distances, which are sensitive to compositional changes in lineages of intermediate abundance (107). Generalized UniFrac distances ($\alpha = 0.5$) were computed using the GUniFrac package.

To deal with uneven total sequence abundance across samples, sequence count composition was scaled to relative abundance. Rarefaction analysis was conducted on the total data set by measuring ASV richness in assemblages randomly subsampled at each 1,000-sequence step. Taxon accumulation was estimated in a random ordering of lakes using 100 permutations in the package vegan (108). Local diversity indices (richness, Pielou's evenness, Shannon diversity, and Faith's phylogenetic diversity) were calculated from rarefied community data (i.e. randomly subsampled to the lowest sample abundance equaling 10,069 sequences) in the R packages vegan (108) and picante (109). PCAs were computed on Hellinger-transformed community data in vegan.

LCBD and β-diversity partitioning analyses were performed on β-diversity estimated using 100 permutations from Hellinger-transformed community data in the package adespatial (110). RV coefficients were computed from the first two principal components or coordinates in the package FactoMineR (111). Nonlinear relationships between β-diversity and untransformed environmental gradients were modeled in GDMs (112, 113) in the package gdm (114). To create GDM site-pair tables, pairwise dissimilarities between sites were weighted proportionally to the total number of sequences associated with each sample. Variable selection for GDMs was performed using backward elimination with 100 permutations per step. To assess the dispersion in taxonomic composition among lakes of the same trophic state, the

mean Bray-Curtis distances of assemblages to trophic state medians (i.e. centroids) calculated across the first two principal coordinates were compared using Tukey's tests of ANOVAs.

Data wrangling and statistical analysis were performed in R v. 4.0.2 (115).

Data availability

Sequence data have been deposited in the European Nucleotide Archive under study accession PRJEB42538 (www.ebi.ac.uk). Scripts are accessible from https://github.com/rebeccagarner/lakepulse_protists.

Acknowledgements

This research was funded by the NSERC Canadian Lake Pulse Network (Strategic Network Grant NETGP-479720) and Canada Research Chairs held by DAW, IGE, and YH. REG and VEO were supported by the NSERC CREATE ÉcoLac training program in lake and fluvial ecology. REG also acknowledges support from a *Fonds de recherche du Québec – Nature et technologies* doctoral research scholarship and the Stephen Bronfman Scholarship in Environmental Studies.

We thank the LakePulse sampling crews and the many landowners, municipal and park employees, lake associations, and First Nations who welcomed and facilitated the sampling effort. We also thank the researchers in LakePulse who contributed to the data set, Zofia Taranu for helpful discussions on statistical methods, and Paul MacKeigan, Cindy Paquette, Marieke Beaulieu, and Katherine Griffiths for their collaboration on data quality assessment and curation.

References

1. Simpson AGB, Eglit Y. 2016. Protist Diversification, p 344–360. *In* Kliman, RM (ed), Encyclopedia of Evolutionary Biology. Academic Press, Waltham, MA.
2. Gadd GM, Raven JA. 2010. Geomicrobiology of Eukaryotic Microorganisms. Geomicrobiol J 27:491–519.
3. Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. 2015. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. Science 347:1257594.

4. Caron DA, Alexander H, Allen AE, Archibald JM, Armbrust EV, Bachy C, Bell CJ, Bharti A, Dyhrman ST, Guida SM, Heidelberg KB, Kaye JZ, Metzner J, Smith SR, Worden AZ. 2017. Probing the evolution, ecology and physiology of marine protists using transcriptomics. Nat Rev Microbiol 15:6–20.

5. Messager ML, Lehner B, Grill G, Nedeva I, Schmitt O. 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nat Commun 7:13603.

6. Tranvik LJ, Downing JA, Cotner JB, Loiselle SA, Striegl RG, Ballatore TJ, Dillon P, Finlay K, Fortino K, Knoll LB, Kortelainen PL, Kutser T, Larsen S, Laurion I, Leech DM, Leigh McCallister S, McKnight DM, Melack JM, Overholt E, Porter JA, Prairie Y, Renwick WH, Roland F, Sherman BS, Schindler DW, Sobek S, Tremblay A, Vanni MJ, Verschoor AM, Von Wachenfeldt E, Weyhenmeyer GA. 2009. Lakes and reservoirs as regulators of carbon cycling and climate. Limnol Oceanogr 54:2298–2314.

7. Beaulieu JJ, DelSontro T, Downing JA. 2019. Eutrophication will increase methane emissions from lakes and impoundments during the 21st century. Nat Commun 10:1375.

8. Anderson NJ, Heathcote AJ, Engstrom DR. 2020. Anthropogenic alteration of nutrient supply increases the global freshwater carbon sink. Sci Adv 6:eaaw2145.

9. Heino J, Alahuhta J, Bini LM, Cai Y, Heiskanen A, Hellsten S, Kortelainen P, Kotamäki N, Tolonen KT, Vihervaara P, Vilmi A, Angeler DG. 2021. Lakes in the era of global change: moving beyond single-lake thinking in maintaining biodiversity and ecosystem services. Biol Rev 96:89–106.

10. Kratz TK, MacIntyre S, Webster KE. 2005. Causes and Consequences of Spatial Heterogeneity in Lakes, p 329–347. *In* Lovett, GM, Jones, C, Turner, MG, Weathers, KC (eds), Ecosystem Function in Heterogeneous Landscapes. Springer, New York, NY.

11. Williamson CE, Dodds W, Kratz TK, Palmer MA. 2008. Lakes and streams as sentinels of environmental change in terrestrial and atmospheric processes. Front Ecol Environ 6:247–254.

12. Strayer DL, Dudgeon D. 2010. Freshwater biodiversity conservation: recent progress and future challenges. J N Am Benthol Soc 29:344–358.

13. Singer D, Seppey CVW, Lentendu G, Dunthorn M, Bass D, Belbahri L, Blandenier Q, Debroas D, de Groot GA, de Vargas C, Domaizon I, Duckert C, Izaguirre I, Koenig I, Mataloni G, Schiaffino MR, Mitchell EAD, Geisen S, Lara E. 2021. Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. Environ Int 146:106262.

14. Smith VH, Schindler DW. 2009. Eutrophication science: where do we go from here? Trends Ecol Evol 24:201–207.

15. O'Reilly CM, Sharma S, Gray DK, Hampton SE, Read JS, Rowley RJ, Schneider P, Lenters JD, McIntyre PB, Kraemer BM, Weyhenmeyer GA, Straile D, Dong B, Adrian R, Allan MG, Anneville O, Arvola L, Austin J, Bailey JL, Baron JS, Brookes JD, Eyto E, Dokulil MT, Hamilton DP, Havens K, Hetherington AL, Higgins SN, Hook S, Izmest'eva LR, Joehnk KD, Kangur K, Kasprzak P, Kumagai M, Kuusisto E, Leshkevich G, Livingstone DM, MacIntyre S, May L, Melack JM, Mueller-Navarra DC, Naumenko M, Noges P, Noges T, North RP, Plisnier P, Rigosi A, Rimmer A, Rogora M, Rudstam LG, Rusak JA, Salmaso N, Samal NR, Schindler DE, Schladow SG, Schmid M, Schmidt SR, Silow E, Soylu ME, Teubner K, Verburg P, Voutilainen A, Watkinson A, Williamson CE, Zhang G. 2015. Rapid and highly variable warming of lake surface waters around the globe. Geophys Res Lett 42:10773–10781.

16. Jane SF, Hansen GJA, Kraemer BM, Leavitt PR, Mincer JL, North RL, Pilla RM, Stetler JT, Williamson CE, Woolway RI, Arvola L, Chandra S, DeGasperi CL, Diemer L, Dunalska J, Erina O, Flaim G, Grossart H-P, Hambright KD, Hein C, Hejzlar J, Janus LL, Jenny J-P, Jones JR, Knoll LB, Leoni B, Mackay E, Matsuzaki S-IS, McBride C, Müller-Navarra DC, Paterson AM, Pierson D, Rogora M, Rusak JA, Sadro S, Saulnier-Talbot E, Schmid M, Sommaruga R, Thiery W, Verburg P, Weathers KC, Weyhenmeyer GA, Yokota K, Rose KC. 2021. Widespread deoxygenation of temperate lakes. Nature 594:66–70.

17. Dugan HA, Skaff NK, Doubek JP, Bartlett SL, Burke SM, Krivak-Tetley FE, Summers JC, Hanson PC, Weathers KC. 2020. Lakes at Risk of Chloride Contamination. Environ Sci Technol 54:6639–6650.

18. Reid AJ, Carlson AK, Creed IF, Eliason EJ, Gell PA, Johnson PTJ, Kidd KA, MacCormack TJ, Olden JD, Ormerod SJ, Smol JP, Taylor WW, Tockner K, Vermaire JC, Dudgeon D, Cooke SJ.

2018. Emerging threats and persistent conservation challenges for freshwater biodiversity. Biol Rev 94:849–873.

19. Capo E, Debroas D, Arnaud F, Perga ME, Chardon C, Domaizon I. 2017. Tracking a century of changes in microbial eukaryotic diversity in lakes driven by nutrient enrichment and climate warming. Environ Microbiol 19:2873–2892.

20. Keck F, Millet L, Debroas D, Etienne D, Galop D, Rius D, Domaizon I. 2020. Assessing the response of micro-eukaryotic diversity to the Great Acceleration using lake sedimentary DNA. Nat Commun 11:3831.

21. Salmaso N, Tolotti M. 2021. Phytoplankton and anthropogenic changes in pelagic environments. Hydrobiologia 848:251–284.

22. Massana R, Castresana J, Balagué V, Guillou L, Romari K, Groisillier A, Valentin K, Pedrós-Alió C. 2004. Phylogenetic and Ecological Analysis of Novel Marine Stramenopiles. Appl Environ Microbiol 70:3528–3534.

23. de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury J-M, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horak A, Jaillon O, Lima-Mendez G, Luke J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E, Boss E, Follows M, Karp-Boss L, Krzic U, Reynaud EG, Sardet C, Sullivan MB, Velayoudon D. 2015. Eukaryotic plankton diversity in the sunlit ocean. Science 348:1261605.

24. Mahé F, de Vargas C, Bass D, Czech L, Stamatakis A, Lara E, Singer D, Mayor J, Bunge J, Sernaker S, Siemensmeyer T, Trautmann I, Romac S, Berney C, Kozlov A, Mitchell EAD, Seppey CVW, Egge E, Lentendu G, Wirth R, Trueba G, Dunthorn M. 2017. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. Nat Ecol Evol 1:0091.

25. Giner CR, Balagué V, Krabberød AK, Ferrera I, Reñé A, Garcés E, Gasol JM, Logares R, Massana R. 2019. Quantifying long-term recurrence in planktonic microbial eukaryotes. Mol Ecol 28:923–935.

26. Logares R, Deutschmann IM, Junger PC, Giner CR, Krabberød AK, Schmidt TSB, Rubinat-Ripoll L, Mestre M, Salazar G, Ruiz-González C, Sebastián M, de Vargas C, Acinas SG, Duarte CM, Gasol JM, Massana R. 2020. Disentangling the mechanisms shaping the surface ocean microbiota. Microbiome 8:55.

27. Vass M, Székely AJ, Lindström ES, Langenheder S. 2020. Using null models to compare bacterial and microeukaryotic metacommunity assembly under shifting environmental conditions. Sci Rep 10:2455.

28. Oliverio AM, Geisen S, Delgado-Baquerizo M, Maestre FT, Turner BL, Fierer N. 2020. The global-scale distributions of soil protists and their contributions to belowground systems. Sci Adv 6:eaax8787.

29. Azam F, Fenchel T, Field J, Gray J, Meyer-Reil L, Thingstad F. 1983. The Ecological Role of Water-Column Microbes in the Sea. Mar Ecol Prog Ser 10:257–263.

30. Sherr EB, Sherr BF. 2002. Significance of predation by protists in aquatic microbial food webs. Antonie Van Leeuwenhoek 81:293–308.

31. Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, Chaffron S, Ignacio-Espinosa JC, Roux S, Vincent F, Bittner L, Darzi Y, Wang J, Audic S, Berline L, Bontempi G, Cabello AM, Coppola L, Cornejo-Castillo FM, D'Ovidio F, De Meester L, Ferrera I, Garet-Delmas M-J, Guidi L, Lara E, Pesant S, Royo-Llonch M, Salazar G, Sanchez P, Sebastian M, Souffreau C, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Gorsky G, Not F, Ogata H, Speich S, Stemmann L, Weissenbach J, Wincker P, Acinas SG, Sunagawa S, Bork P, Sullivan MB, Karsenti E, Bowler C, de Vargas C, Raes J. 2015. Determinants of community structure in the global plankton interactome. Science 348:1262073.

32. Bjorbækmo MFM, Evenstad A, Røsæg LL, Krabberød AK, Logares R. 2020. The planktonic protist interactome: where do we stand after a century of research? ISME J 14:544–559.

666 33. Piwosz K, Mukherjee I, Salcher MM, Grujčić V, Šimek K. 2021. CARD-FISH in the Sequencing
667 Era: Opening a New Universe of Protistan Ecology. Front Microbiol 12:640066.
668 34. Gast RJ, Sanders RW, Caron DA. 2009. Ecological strategies of protists and their symbiotic
669 relationships with prokaryotic microbes. Trends Microbiol 17:563–569.
670 35. Monteil CL, Vallenet D, Menguy N, Benzerara K, Barbe V, Fouteau S, Cruaud C, Floriani M,
671 Viollier E, Adryanczyk G, Leonhardt N, Faivre D, Pignol D, López-García P, Weld RJ, Lefevre CT.
672 2019. Ectosymbiotic bacteria at the origin of magnetoreception in a marine protist. Nat Microbiol
673 4:1088–1095.
674 36. Husnik F, Tashyreva D, Boscaro V, George EE, Lukeš J, Keeling PJ. 2021. Bacterial and
675 archaeal symbioses with protists. Curr Biol 31:R862–R877.
676 37. Coy SR, Gann ER, Pound HL, Short SM, Wilhelm SW. 2018. Viruses of Eukaryotic Algae:
677 Diversity, Methods for Detection, and Future Directions. Viruses 10:487.
678 38. Ramond P, Sourisseau M, Simon N, Romac S, Schmitt S, Rigaut-Jalabert F, Henry N, de Vargas
679 C, Siano R. 2019. Coupling between taxonomic and functional diversity in protistan coastal
680 communities. Environ Microbiol 21:730–749.
681 39. Payne RJ. 2013. Seven Reasons Why Protists Make Useful Bioindicators. Acta Protozool
682 52:105–113.
683 40. Sagova-Mareckova M, Boenigk J, Bouchez A, Cermakova K, Chonova T, Cordier T, Eisendle U,
684 Elersek T, Fazi S, Fleituch T, Frühe L, Gajdosova M, Graupner N, Haegerbaeumer A, Kelly A-M,
685 Kopecky J, Leese F, Nõges P, Orlic S, Panksep K, Pawlowski J, Petrusek A, Piggott JJ, Rusch
686 JC, Salis R, Schenk J, Simek K, Stovicek A, Strand DA, Vasquez MI, Vrålstad T, Zlatkovic S,
687 Zupancic M, Stoeck T. 2021. Expanding ecological assessment by integrating microorganisms
688 into routine freshwater biomonitoring. Water Res 191:116767.
689 41. Filker S, Sommaruga R, Vila I, Stoeck T. 2016. Microbial eukaryote plankton communities of high-
690 mountain lakes from three continents exhibit strong biogeographic patterns. Mol Ecol 25:2286–
691 2301.
692 42. Khomich M, Kauserud H, Logares R, Rasconi S, Andersen T. 2017. Planktonic protistan
693 communities in lakes along a large-scale environmental gradient. FEMS Microbiol Ecol 93:fiw231.

694 43. Boenigk J, Wodniok S, Bock C, Beisser D, Hempel C, Grossmann L, Lange A, Jensen M. 2018.
695 Geographic distance and mountain ranges structure freshwater protist communities on a
696 European scale. Metabarcoding and Metagenomics 2:e21519.
697 44. Bock C, Jensen M, Forster D, Marks S, Nuy J, Psenner R, Beisser D, Boenigk J. 2020. Factors
698 shaping community patterns of protists and bacteria on a European scale. Environ Microbiol
699 22:2243–2260.
700 45. Ptacnik R, Lepistö L, Willén E, Brettum P, Andersen T, Rekolainen S, Lyche Solheim A, Carvalho
701 L. 2008. Quantitative responses of lake phytoplankton to eutrophication in Northern Europe.
702 Aquat Ecol 42:227–236.
703 46. Stomp M, Huisman J, Mittelbach GG, Litchman E, Klausmeier CA. 2011. Large-scale biodiversity
704 patterns in freshwater phytoplankton. Ecology 92:2096–2107.
705 47. Maileht K, Nõges T, Nõges P, Ott I, Mischke U, Carvalho L, Dudley B. 2013. Water colour,
706 phosphorus and alkalinity are the major determinants of the dominant phytoplankton species in
707 European lakes. Hydrobiologia 704:115–126.
708 48. Keeling PJ, del Campo J. 2017. Marine Protists Are Not Just Big Bacteria. Curr Biol 27:R541–
709 R549.
710 49. Stoecker DK, Hansen PJ, Caron DA, Mitra A. 2017. Mixotrophy in the Marine Plankton. Ann Rev
711 Mar Sci 9:311–335.
712 50. Berninger U-G, Caron DA, Sanders RW. 1992. Mixotrophic algae in three ice-covered lakes of the
713 Pocono Mountains, U.S.A. Freshw Biol 28:263–272.
714 51. Hitchman RB, Jones HLJ. 2000. The role of mixotrophic protists in the population dynamics of the
715 microbial food web in a small artificial pond. Freshw Biol 43:231–241.
716 52. Huot Y, Brown CA, Potvin G, Antoniades D, Baulch HM, Beisner BE, Bélanger S, Brazeau S,
717 Cabana H, Cardille JA, del Giorgio PA, Gregory-Eaves I, Fortin M-J, Lang AS, Laurion I,
718 Maranger R, Prairie YT, Rusak JA, Segura PA, Siron R, Smol JP, Vinebrooke RD, Walsh DA.
719 2019. The NSERC Canadian Lake Pulse Network: A national assessment of lake health providing
720 science for water management in a changing climate. Sci Total Environ 695:133668.

721 53. Minns CK, Moore JE, Shuter BJ, Mandrak NE. 2008. A preliminary national analysis of some key
722 characteristics of Canadian lakes. Can J Fish Aquat Sci 65:1763–1778.
723 54. Giner CR, Pernice MC, Balagué V, Duarte CM, Gasol JM, Logares R, Massana R. 2020. Marked
724 changes in diversity and relative activity of picoeukaryotes with depth in the world ocean. ISME J
725 14:437–449.
726 55. Ortiz-Álvarez R, Triadó-Margarit X, Camarero L, Casamayor EO, Catalan J. 2018. High
727 planktonic diversity in mountain lakes contains similar contributions of autotrophic, heterotrophic
728 and parasitic eukaryotic life forms. Sci Rep 8:4457.
729 56. Massana R, Gobet A, Audic S, Bass D, Bittner L, Boutte C, Chambouvet A, Christen R, Claverie
730 J-M, Decelle J, Dolan JR, Dunthorn M, Edvardsen B, Forn I, Forster D, Guillou L, Jaillon O,
731 Kooistra WHCF, Logares R, Mahé F, Not F, Ogata H, Pawlowski J, Pernice MC, Probert I, Romac
732 S, Richards T, Santini S, Shalchian-Tabrizi K, Siano R, Simon N, Stoeck T, Vaulot D, Zingone A,
733 de Vargas C. 2015. Marine protist diversity in European coastal waters and sediments as
734 revealed by high-throughput sequencing. Environ Microbiol 17:4035–4049.
735 57. Parfrey LW, Walters WA, Lauber CL, Clemente JC, Berg-Lyons D, Teiling C, Kodira C, Mohiuddin
736 M, Brunelle J, Driscoll M, Fierer N, Gilbert JA, Knight R. 2014. Communities of microbial
737 eukaryotes in the mammalian gut within the context of environmental eukaryotic diversity. Front
738 Microbiol 5:298.
739 58. Debroas D, Domaizon I, Humbert J-FF, Jardillier L, Lepère C, Oudart A, Taïb N, Lepére C,
740 Oudart A, Taib N. 2017. Overview of freshwater microbial eukaryotes diversity: a first analysis of
741 publicly available metabarcoding data. FEMS Microbiol Ecol 93:fix023.
742 59. Grujcic V, Nuy JK, Salcher MM, Shabarova T, Kasalicky V, Boenigk J, Jensen M, Simek K. 2018.
743 Cryptophyta as major bacterivores in freshwater summer plankton. ISME J 12:1668–1681.
744 60. Ibarbalz FM, Henry N, Brandão MC, Martini S, Busseni G, Byrne H, Coelho LP, Endo H, Gasol
745 JM, Gregory AC, Mahé F, Rigonato J, Royo-Llonch M, Salazar G, Sanz-Sáez I, Scalco E,
746 Soviadan D, Zayed AA, Zingone A, Labadie K, Ferland J, Marec C, Kandels S, Picheral M, Dimier
747 C, Poulain J, Pisarev S, Carmichael M, Pesant S, Acinas SG, Babin M, Bork P, Boss E, Bowler C,
748 Cochrane G, de Vargas C, Follows M, Gorsky G, Grimsley N, Guidi L, Hingamp P, Iudicone D,

Jaillon O, Karp-Boss L, Karsenti E, Not F, Ogata H, Poulton N, Raes J, Sardet C, Speich S, Stemmann L, Sullivan MB, Sunagawa S, Wincker P, Pelletier E, Bopp L, Lombard F, Zinger L. 2019. Global trends in marine plankton diversity across kingdoms of life. Cell 179:1084–1097.

61. Martin SL, Soranno PA. 2006. Lake landscape position: Relationships to hydrologic connectivity and landscape features. Limnol Oceanogr 51:801–814.

62. Paquette C, Gregory-Eaves I, Beisner BE. 2021. Multi-scale biodiversity analyses identify the importance of continental watersheds in shaping lake zooplankton biogeography. J Biogeogr 48:2298–2311.

63. Jaworski GHM, Talling JF, Heaney SI. 2003. Potassium dependence and phytoplankton ecology: an experimental study. Freshw Biol 48:833–840.

64. Wilken S, Choi CJ, Worden AZ. 2020. Contrasting mixotrophic lifestyles reveal different ecological niches in two closely related marine protists. J Phycol 56:52–67.

65. Katechakis A, Stibor H. 2006. The mixotroph *Ochromonas tuberculata* may invade and suppress specialist phago- and phototroph plankton communities depending on nutrient conditions. Oecologia 148:692–701.

66. Saad JF, Unrein F, Tribelli PM, López N, Izaguirre I. 2016. Influence of lake trophic conditions on the dominant mixotrophic algal assemblages. J Plankton Res 38:818–829.

67. Wilken S, Soares M, Urrutia-Cordero P, Ratcovich J, Ekvall MK, Van Donk E, Hansson L-A. 2018. Primary producers or consumers? Increasing phytoplankton bacterivory along a gradient of lake warming and browning. Limnol Oceanogr 63:S142–S155.

68. Fischer R, Giebel H-A, Hillebrand H, Ptacnik R. 2017. Importance of mixotrophic bacterivory can be predicted by light and loss rates. Oikos 126:713–722.

69. Hansson TH, Grossart H, Giorgio PA, St-Gelais NF, Beisner BE. 2019. Environmental drivers of mixotrophs in boreal lakes. Limnol Oceanogr 64:1688–1705.

70. Schindler DW, Armstrong FAJ, Holmgren SK, Brunskill GJ. 1971. Eutrophication of Lake 227, Experimental Lakes Area, Northwestern Ontario, by Addition of Phosphate and Nitrate. J Fish Res Board Canada 28:1763–1782.

776 71. Liang Z, Soranno PA, Wagner T. 2020. The role of phosphorus and nitrogen on chlorophyll *a*:
777 Evidence from hundreds of lakes. Water Res 185:116236.
778 72. MacKeigan PW, Garner RE, Monchamp M-È, Walsh DA, Onana VE, Kraemer SA, Pick FR,
779 Beisner BE, Agbeti MD, da Costa NB, Shapiro BJ, Gregory-Eaves I. 2022. Comparing
780 microscopy and DNA metabarcoding techniques for identifying cyanobacteria assemblages
781 across hundreds of lakes. Harmful Algae 113:102187.
782 73. Maloufi S, Catherine A, Mouillot D, Louvard C, Couté A, Bernard C, Troussellier M. 2016.
783 Environmental heterogeneity among lakes promotes hyper β-diversity across phytoplankton
784 communities. Freshw Biol 61:633–645.
785 74. Monchamp M-E, Spaak P, Domaizon I, Dubois N, Bouffard D, Pomati F. 2018. Homogenization of
786 lake cyanobacterial communities over a century of climate change and eutrophication. Nat Ecol
787 Evol 2:317–324.
788 75. Aguilar P, Sommaruga R. 2020. The balance between deterministic and stochastic processes in
789 structuring lake bacterioplankton community over time. Mol Ecol 29:3117–3130.
790 76. Baud A, Jenny J-P, Francus P, Gregory-Eaves I. 2021. Global acceleration of lake sediment
791 accumulation rates associated with recent human population growth and land-use changes. J
792 Paleolimnol 66:453–467.
793 77. Mitchell P, Prepas EE (eds). 1990. Atlas of Alberta Lakes. University of Alberta Press, Edmonton,
794 AB.
795 78. Dornelas M, Gotelli NJ, McGill B, Shimadzu H, Moyes F, Sievers C, Magurran AE. 2014.
796 Assemblage Time Series Reveal Biodiversity Change but Not Systematic Loss. Science
797 344:296–299.
798 79. Dillon PJ, Jeffries DS, Snyder W, Reid R, Yan ND, Evans D, Moss J, Scheider WA. 1978. Acidic
799 Precipitation in South-Central Ontario: Recent Observations. J Fish Res Board Canada 35:809–
800 815.
801 80. Delgado A, Gómez JA. 2016. The Soil. Physical, Chemical and Biological Properties, p 15–26. *In*
802 Villalobos, FJ, Fereres, E (eds), Principles of Agronomy for Sustainable Agriculture. Springer
803 International Publishing, Cham.

804 81. Fierer N. 2017. Embracing the unknown: disentangling the complexities of the soil microbiome.
805 Nat Rev Microbiol 15:579–590.
806 82. Nearing MA, Pruski FF, O'Neal MR. 2004. Expected climate change impacts on soil erosion
807 rates: A review. J Soil Water Conserv 59:43–50.
808 83. Jeppesen E, Kronvang B, Meerhoff M, Søndergaard M, Hansen KM, Andersen HE, Lauridsen TL,
809 Liboriussen L, Beklioglu M, Özen A, Olesen JE. 2009. Climate Change Effects on Runoff,
810 Catchment Phosphorus Loading and Lake Ecological State, and Potential Adaptations. J Environ
811 Qual 38:1930–1941.
812 84. Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, Chapin FS, Coe MT, Daily
813 GC, Gibbs HK, Helkowski JH, Holloway T, Howard EA, Kucharik CJ, Monfreda C, Patz JA,
814 Prentice IC, Ramankutty N, Snyder PK. 2005. Global Consequences of Land Use. Science
815 309:570–574.
816 85. Barak P, Jobe BO, Krueger AR, Peterson LA, Laird DA. 1997. Effects of long-term soil
817 acidification due to nitrogen fertilizer inputs in Wisconsin. Plant Soil 197:61–69.
818 86. Ontl TA, Schulte LA. 2012. Soil Carbon Storage. Nat Educ Knowl 3:35.
819 87. Kaushal SS, Mayer PM, Vidon PG, Smith RM, Pennino MJ, Newcomer TA, Duan S, Welty C, Belt
820 KT. 2014. Land use and climate variability amplify carbon, nutrient, and contaminant pulses: a
821 review with management implications. J Am Water Resour Assoc 50:585–614.
822 88. Oliva A, Garner RE, Walsh D, Huot Y. 2022. The occurrence of potentially pathogenic fungi and
823 protists in Canadian lakes predicted using geomatics, *in situ* and satellite-derived variables:
824 Towards a tele-epidemiological approach. Water Res 209:117935.
825 89. Wiken EB, Gauthier D, Marshall I, Lawton K, Hirvonen H. 1996. A Perspective on Canada's
826 Ecosystems: An Overview of the Terrestrial and Marine Ecozones. *In* Canadian Council on
827 Ecological Areas Occasional Papers. Ottawa, ON.
828 90. NSERC Canadian Lake Pulse Network. 2021. NSERC Canadian Lake Pulse Network field
829 manual 2017 - 2018 - 2019 surveys. Varin MP, Beaulieu ML, Huot Y (eds). Université de
830 Sherbrooke.

831 91. Gast RJ, Dennett MR, Caron DA. 2004. Characterization of Protistan Assemblages in the Ross
832 Sea, Antarctica, by Denaturing Gradient Gel Electrophoresis. Appl Environ Microbiol 70:2028–
833 2037.
834 92. Van de Peer Y. 2000. The European Small Subunit Ribosomal RNA database. Nucleic Acids Res
835 28:175–176.
836 93. Capo E, Debroas D, Arnaud F, Guillemot T, Bichet V, Millet L, Gauthier E, Massa C, Develle AL,
837 Pignol C, Lejzerowicz F, Domaizon I. 2016. Long-term dynamics in microbial eukaryotes
838 communities: a palaeolimnological view based on sedimentary DNA. Mol Ecol 25:5925–5943.
839 94. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
840 EMBnet.journal 17:10–12.
841 95. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-
842 resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583.
843 96. Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C,
844 Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara
845 E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J,
846 Probert I, Sauvadet A-L, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R. 2012. The
847 Protist Ribosomal Reference database (PR^2): a catalog of unicellular eukaryote Small Sub-Unit
848 rRNA sequences with curated taxonomy. Nucleic Acids Res 41:D597–D604.
849 97. Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment,
850 interactive sequence choice and visualization. Brief Bioinform 20:1160–1166.
851 98. Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW,
852 Burki F, Cárdenas P, Čepička I, Chistyakova L, Campo J, Dunthorn M, Edvardsen B, Eglit Y,
853 Guillou L, Hampl V, Heiss AA, Hoppenrath M, James TY, Karnkowska A, Karpov S, Kim E,
854 Kolisko M, Kudryavtsev A, Lahr DJG, Lara E, Le Gall L, Lynn DH, Mann DG, Massana R, Mitchell
855 EAD, Morrow C, Park JS, Pawlowski JW, Powell MJ, Richter DJ, Rueckert S, Shadwick L,
856 Shimano S, Spiegel FW, Torruella G, Youssef N, Zlatogursky V, Zhang Q. 2019. Revisions to the
857 Classification, Nomenclature, and Diversity of Eukaryotes. J Eukaryot Microbiol 66:4–119.

858 99. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
859 BLAST+: architecture and applications. BMC Bioinformatics 10:421.
860 100. Canadian Council of Ministers of the Environment. 2004. Canadian water quality guidelines for
861 the protection of aquatic life: Phosphorus: Canadian Guidance Framework for the Management of
862 Freshwater Systems. *In* Canadian environmental quality guidelines. Winnipeg, MB.
863 101. Muñoz Sabater J. 2019. ERA5-Land hourly data from 1981 to present. Copernicus Climate
864 Change Service (C3S) Climate Data Store (CDS).
865 102. Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A,
866 Shangguan W, Wright MN, Geng X, Bauer-Marschallinger B, Guevara MA, Vargas R, MacMillan
867 RA, Batjes NH, Leenaars JGB, Ribeiro E, Wheeler I, Mantel S, Kempen B. 2017. SoilGrids250m:
868 Global gridded soil information based on machine learning. PLoS One 12:e0169748.
869 103. Becker RA, Wilks AR, Brownrigg R, Minka TP, Deckmyn A. 2018. maps: Draw geographical
870 maps. CRAN.
871 104. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: Accurate high-throughput multiple sequence
872 alignment of ribosomal RNA genes. Bioinformatics 28:1823–1829.
873 105. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012.
874 The SILVA ribosomal RNA gene database project: improved data processing and web-based
875 tools. Nucleic Acids Res 41:D590–D596.
876 106. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for
877 Large Alignments. PLoS One 5:e9490.
878 107. Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, Collman RG, Bushman FD, Li H.
879 2012. Associating microbiome composition with environmental covariates using generalized
880 UniFrac distances. Bioinformatics 28:2106–2113.
881 108. Oksanen J, Blanchet FG, Kindt R, Legen- P, Minchin PR, Hara RBO, Simpson GL, Soly- P,
882 Stevens MHH, Wagner H. 2020. vegan: Community ecology package. CRAN.
883 109. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb
884 CO. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–
885 1464.

886 110. Dray S, Bauman D, Blanchet G, Borcard D, Clappe S, Guenard G, Jombart T, Larocque G,
887 Legendre P, Madi N, Wagner HH. 2021. adespatial: Multivariate multiscale spatial analysis.
888 CRAN.
889 111. Lê S, Josse J, Husson F. 2008. FactoMineR: An R Package for Multivariate Analysis. J Stat Softw
890 25:253–258.
891 112. Ferrier S, Manion G, Elith J, Richardson K. 2007. Using generalized dissimilarity modelling to
892 analyse and predict patterns of beta diversity in regional biodiversity assessment. Divers Distrib
893 13:252–264.
894 113. Rosauer DF, Ferrier S, Williams KJ, Manion G, Keogh JS, Laffan SW. 2014. Phylogenetic
895 generalised dissimilarity modelling: a new approach to analysing and predicting spatial turnover in
896 the phylogenetic composition of communities. Ecography 37:21–32.
897 114. Matthew C. Fitzpatrick, Mokany K, Manion G, Lisk M, Ferrier S, Nieto-Lugilde D. 2021. gdm:
898 Generalized dissimilarity modeling. CRAN.
899 115. R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for
900 Statistical Computing, Vienna, Austria.
901

902 **Figure legends**

903 **Figure 1**. Diversity and distributions of protists across 366 Canadian lakes. Lake trophic states and
904 ecozones are shown on the map of sampling sites as coloured circles and polygons, respectively. The
905 relative sequence abundance of protist taxonomic divisions in each lake is represented in the inner tract
906 of bar plots. The middle tract of bar plots shows the relative sequence abundance of trophic modes. The
907 heatmap on the outer edge illustrates the proportions of land use and land cover associated with the
908 watershed of each lake, whose trophic state is represented in an adjacent coloured circle. Watershed
909 land use proportions are hierarchically clustered to highlight the relationship between agriculture and
910 trophic state.

911 **Figure 2**. Accumulation and incidence of genotypes across lakes. A) Accumulation curve of genotypes in
912 a random ordering of lakes. Vertical bars are standard deviations. The accumulation of new genotypes
913 was rapid in the first ~100 lakes, followed by a gradual deceleration. B) Incidence of genotypes across

914 lakes. Most taxa are distributed across one or a few lakes, whereas a few taxa (magnified in inset plot)
915 are widely distributed. ASV taxonomic classifications are coloured according to the taxonomic divisions in
916 the legend of Figure 1.
917 **Figure 3**. Local protist diversity of each lake. A) Rarefied Shannon diversity index calculated for each
918 protist assemblage across the Canadian landscape. Ecozones are identified in the map legend of Figure
919 1. B) Local diversity metrics categorized by lake trophic state. TP concentration thresholds and sample
920 sizes of lakes categorized under each trophic state are: ultraoligotrophic (TP <4 µg/L; *n* = 4), oligotrophic
921 (4 – 10 µg/L; *n* = 48), mesotrophic (10 – 20 µg/L; *n* = 132), mesoeutrophic (20 – 35 µg/L; *n* = 71),
922 eutrophic (35 – 100 µg/L; *n* = 61), and hypereutrophic (>100 µg/L; *n* = 50).
923 **Figure 4**. Taxonomic and phylogenetic variation of protist assemblages among lakes. A) PCA of the
924 taxonomic variation among protist assemblages at the level of individual ASVs. The relative contributions
925 and taxonomic assignments are shown for the top 7 ASVs (an arbitrary cut-off selected for illustrative
926 clarity) contributing to the variation explained by the first two PC dimensions. B) PCoA of the phylogenetic
927 variation among protist assemblages. Ecozones and trophic state classifications of lakes are represented
928 by letter symbols and colours, respectively.
929 **Figure 5**. Local contribution to β-diversity (LCBD) of protist assemblages across the Canadian landscape.
930 LCBD describes the taxonomic uniqueness of a given assemblage, i.e. how much the taxonomic
931 composition differs from the rest of the communities in the landscape. Ecozones are identified in the map
932 legend of Figure 1.
933 **Figure 6**. PCAs of the taxonomic variation among (A) phototroph, (B) heterotroph, and (C) mixotroph
934 assemblages at the level of individual ASVs. The relative contributions and taxonomic assignments are
935 shown for the top 7 ASVs contributing to the variation explained by the first two PC dimensions. Ecozone
936 affiliations and trophic state classifications of lakes are represented by letter symbols and colours,
937 respectively.
938
939 **Tables**
940 **Table 1**. Percent deviance explained by GDMs. GDMs were constructed using various community
941 response data and categories of environmental explanatory variables. Models that were not statistically

942 significant ($p \geq 0.05$) are denoted by NS. *Analysis was performed only on taxonomic composition
943 response data.

Explanatory variables	All protists		Trophic mode*		
	Taxonomy	Phylogeny	Phototrophs	Heterotrophs	Mixotrophs
Physicochemistry	38	33	35	42	18
Watershed	15	16	20	19	NS
Morphometry	6	8	5	NS	NS
Weather	NS	NS	NS	3	NS
Geography	NS	NS	NS	2	NS

944

945 **Supplemental materials**

946 **Figure S1**. Distributions of A) geography, B) lake morphometry, C) watershed, D) weather, and E)
947 physicochemical variables by ecozone. Dashed lines denote ecozone medians.

948 **Figure S2**. Principal component analysis of sites based on lake and watershed environmental
949 characteristics.

950 **Figure S3**. Rarefaction curve showing the number of ASVs detected as a function of sample size across
951 all sites.

952 **Figure S4**. Sequence similarities of ASVs to known 18S rRNA gene V7 region diversity. A) Frequency of
953 ASVs as a function of sequence identity with the top hit reference in PR^2. B) Number of sequences
954 associated with the sequence identities of top hits. Bar colours represent the division-rank taxonomic
955 assignment top hit references in PR^2.

956 **Figure S5**. Map of rarefied ASV richness.

957 **Figure S6**. Partial effects of A) physicochemical, B) watershed, and C) lake morphometry variables on the
958 taxonomic and phylogenetic turnover of protist assemblages. Points along the x-axis denote the
959 distribution of explanatory variables and are coloured according to lake trophic state (see map legend in
960 Figure 1).

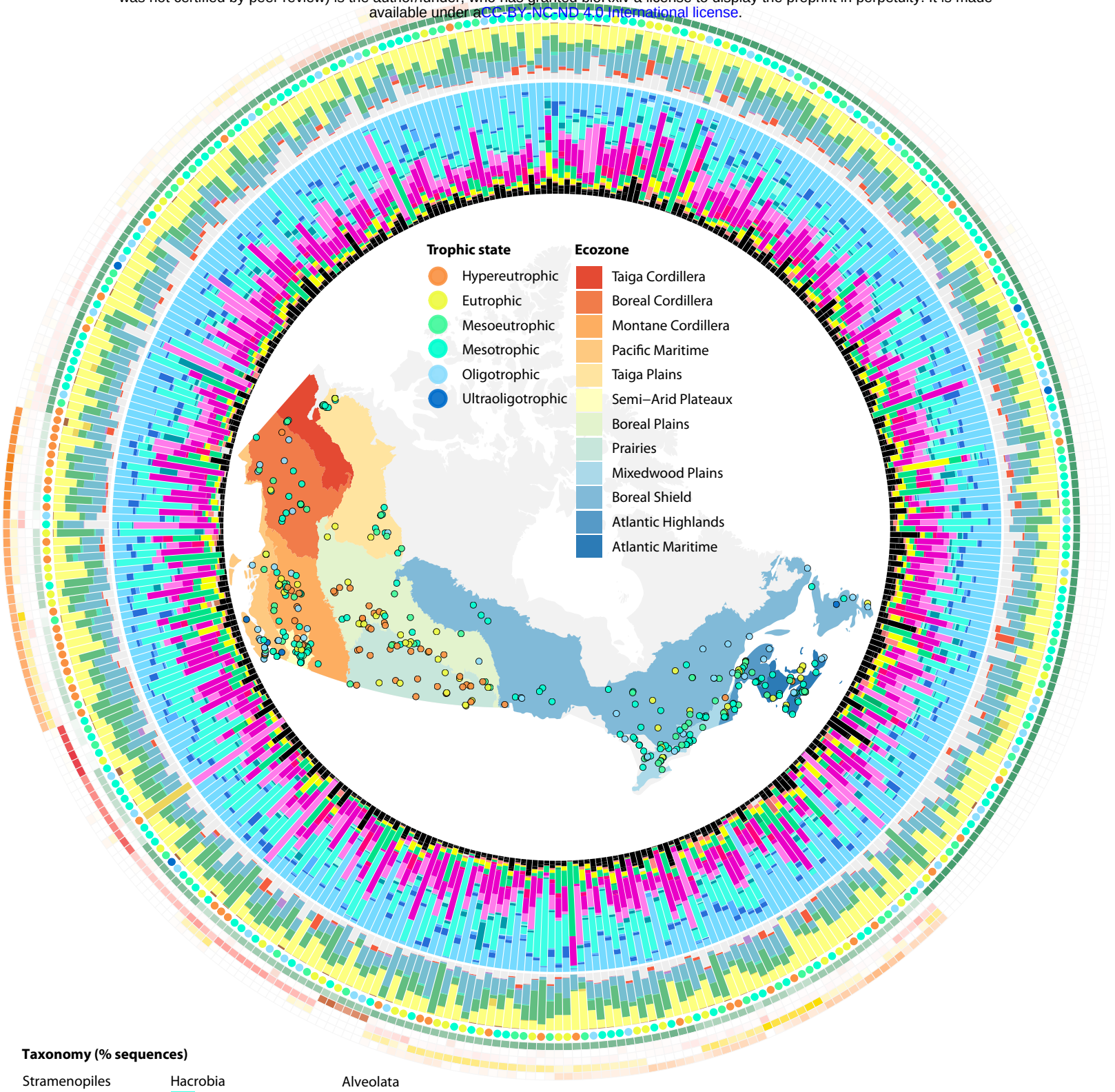
961 **Figure S7**. Phototroph, heterotroph, and mixotroph assemblage distances from trophic state centroids.

962 **Figure S8**. Partial effects of A) physicochemical, B) watershed, C) lake morphometry, D) weather, and E)

963 geography variables on the taxonomic turnover of phototroph, heterotroph, and mixotroph assemblages.
964 Points along the x-axis denote the distribution of explanatory variables and are coloured according to lake
965 trophic state (see map legend in Figure 1).

966 **Table S1**. ASV taxonomic and trophic functional annotations.

967 **Table S2**. Explanatory variables and relative importance of individual predictors in GDMs. Partial effects
968 (i.e. relative importance) of individual predictors within models (parenthesized and presented in
969 descending order of variable importance) were determined as the sum of I-spline coefficients equaling the
970 magnitude of turnover. Non-significant models are denoted NS. *For models of turnover in trophic mode
971 assemblages, analysis was performed only on taxonomic composition response data.



Taxonomy (% sequences)

Stramenopiles
- Ochrophyta
- Opalozoa
- Pseudofungi
- Sagenista

Archaeplastida
- Chlorophyta
- Streptophyta
- Rhodophyta
- Glaucophyta

Rhizaria
- Cercozoa
- Foraminifera

Hacrobia
- Cryptophyta
- Katablepharidophyta
- Centroheliozoa
- Haptophyta
- Telonemia

Opisthokonta
- Choanoflagellida
- Mesomycetozoa

Excavata
- Discoba
- Metamonada
- Malawimonadidae

Alveolata
- Dinoflagellata
- Ciliophora
- Apicomplexa
- Perkinsea

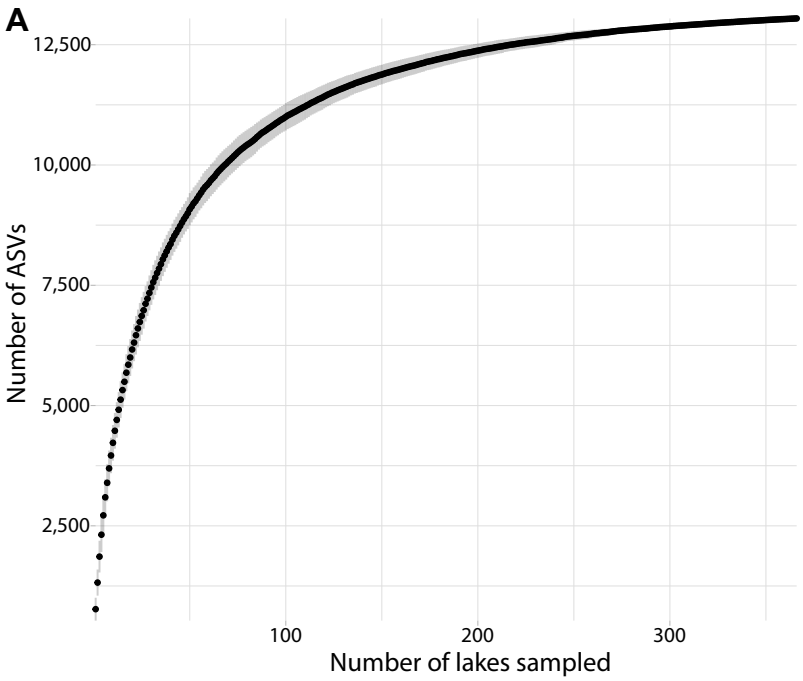
Amoebozoa
- Lobosa
- Conosa
- Breviatea

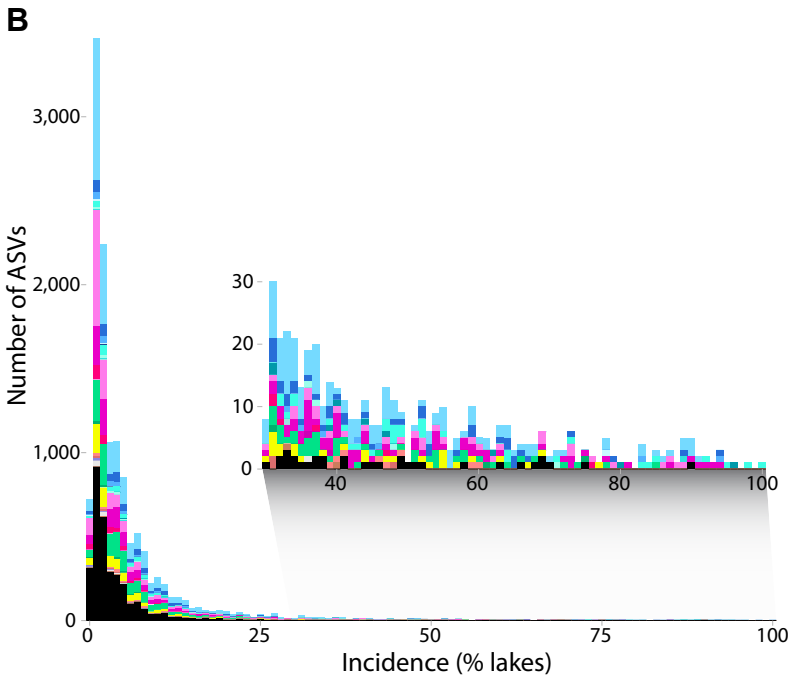
Apusozoa
- Hilomonadea
- Apusomonadidae

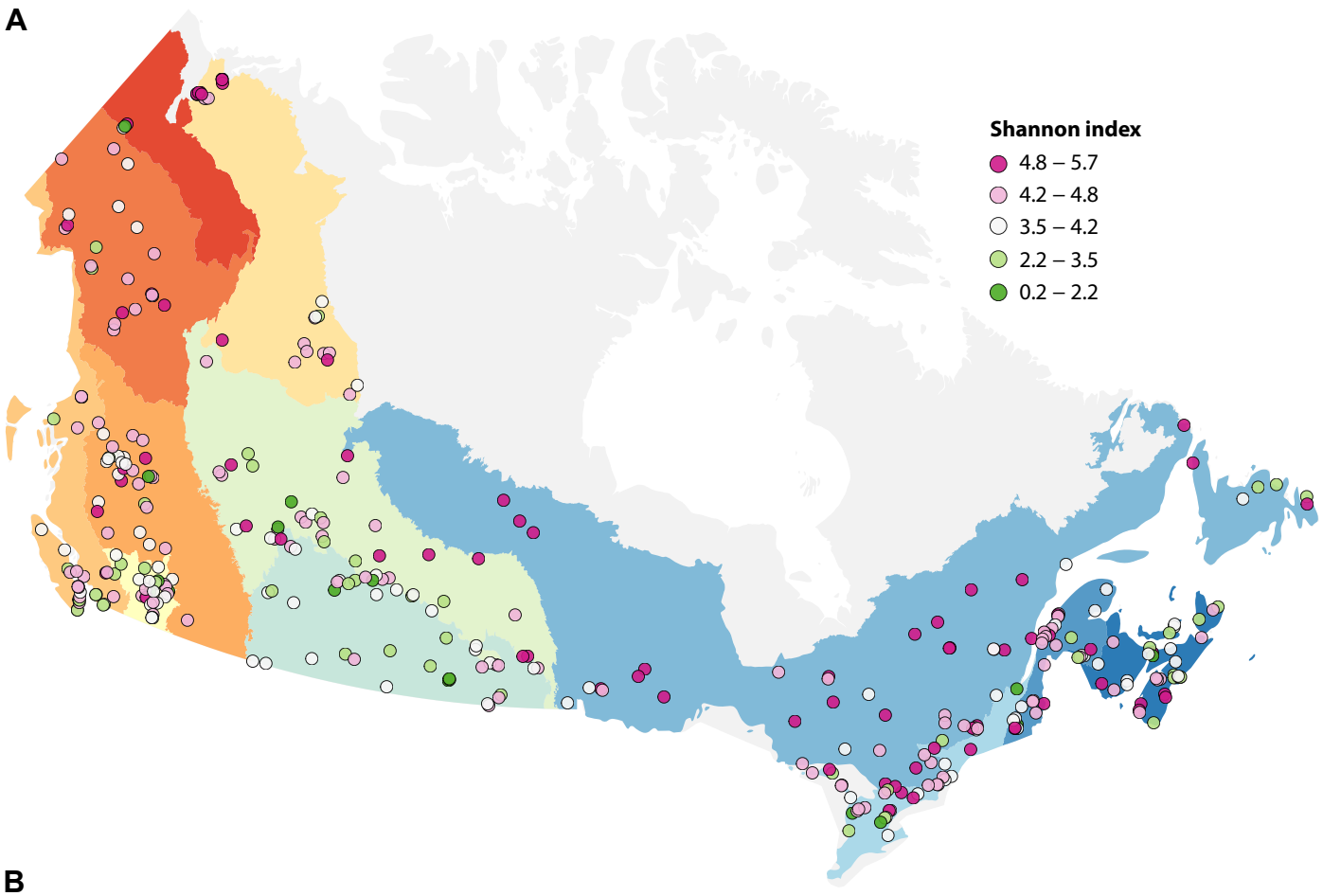
- Unassigned eukaryotes

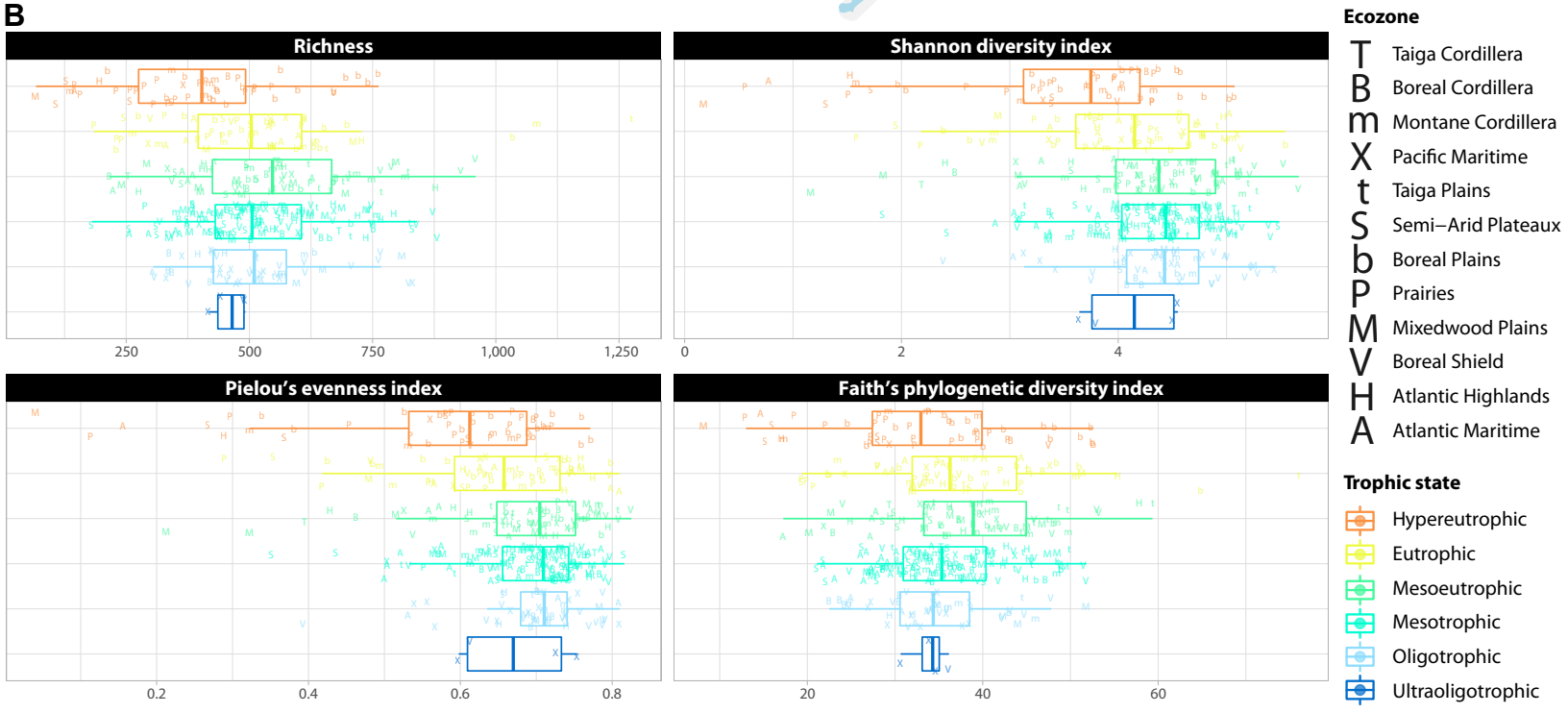
Trophic mode (% sequences)
- Phototrophs, heterotrophs, mixotrophs
- Phototrophs
- Phototrophs, mixotrophs
- Mixotrophs
- Heterotrophs, mixotrophs
- Heterotrophs
- Heterotrophs, parasites
- Parasites
- Unassigned

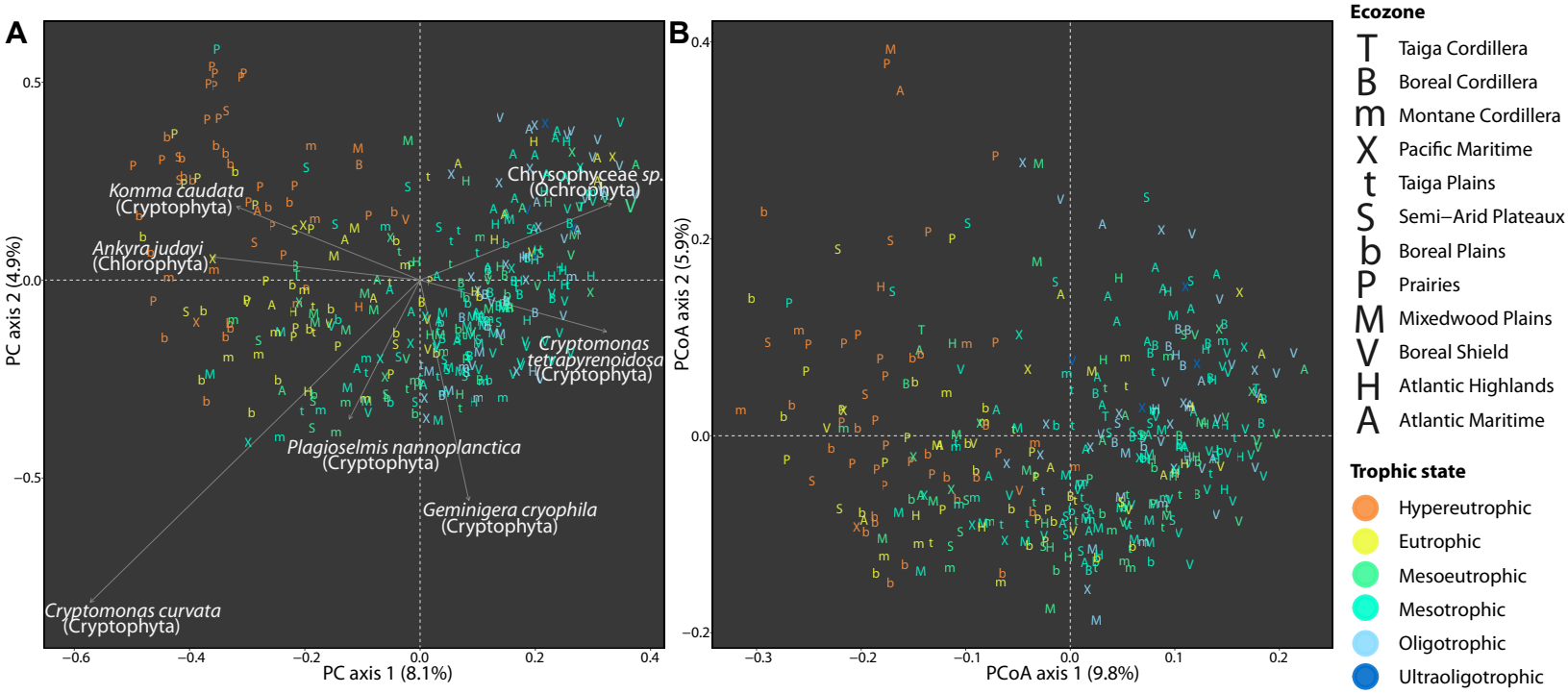
Watershed land use/cover (%)
- 0 – 83 Agriculture
- 0 – 45 Pasture
- 0 – 96 Built
- 0 – 59 Forestry
- 0 – 100 Natural landscapes

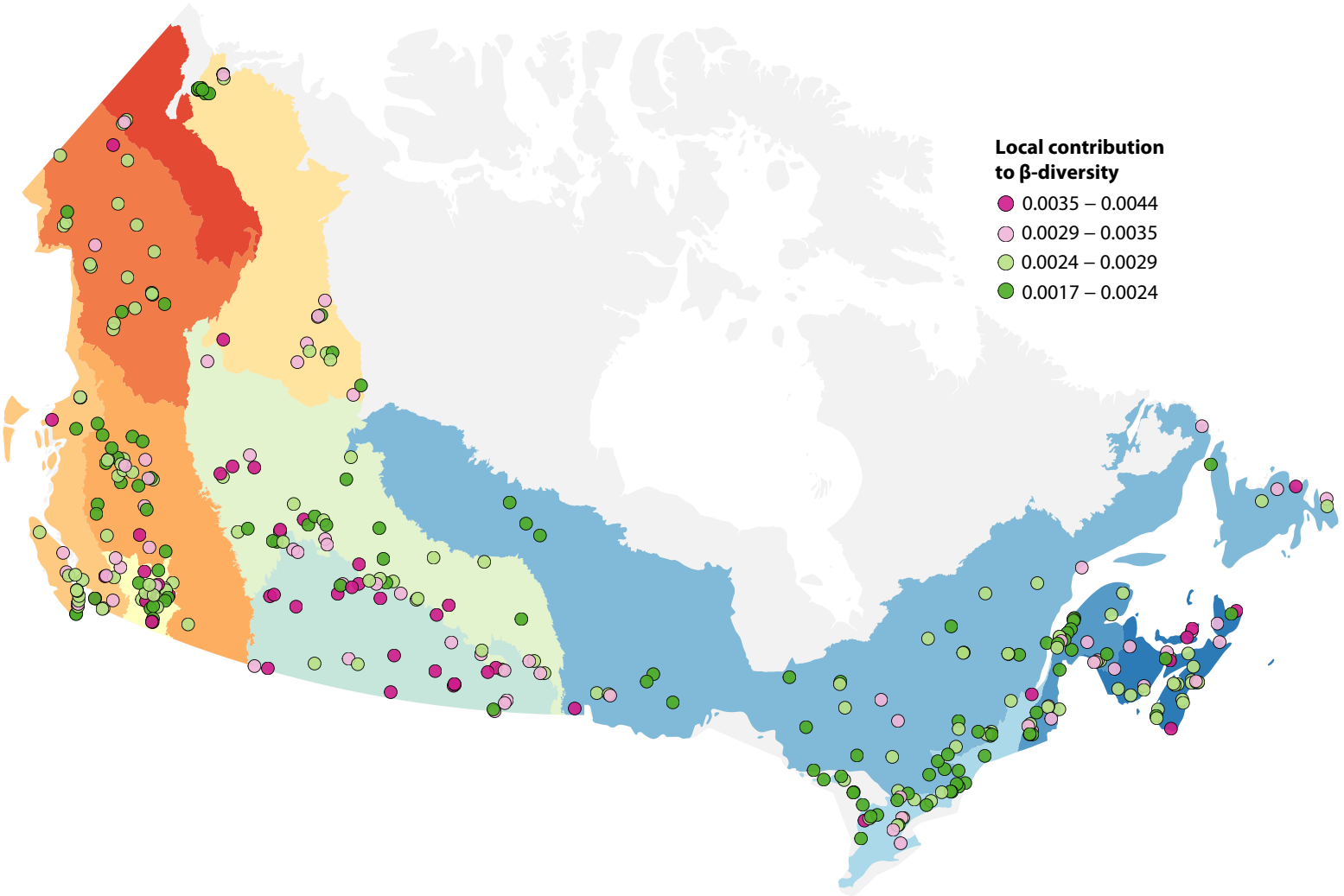


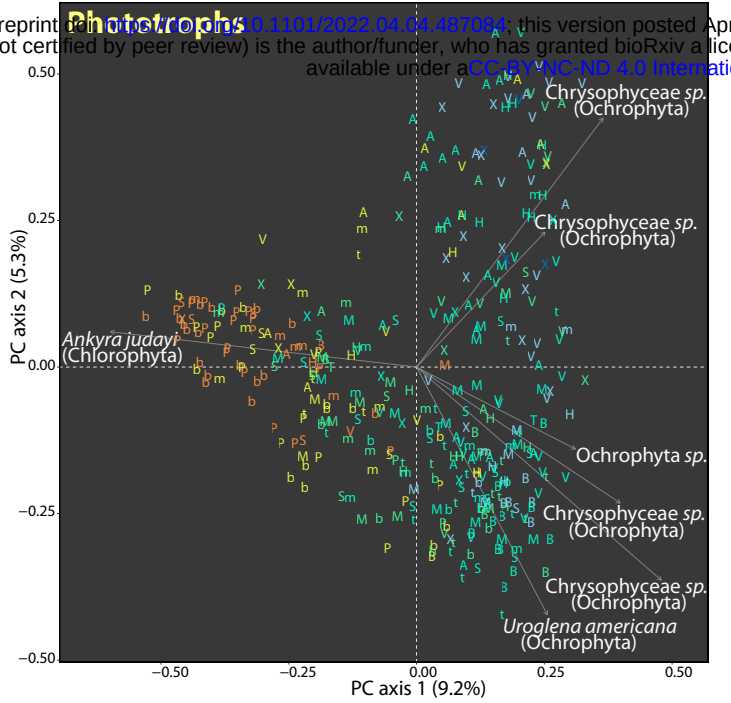


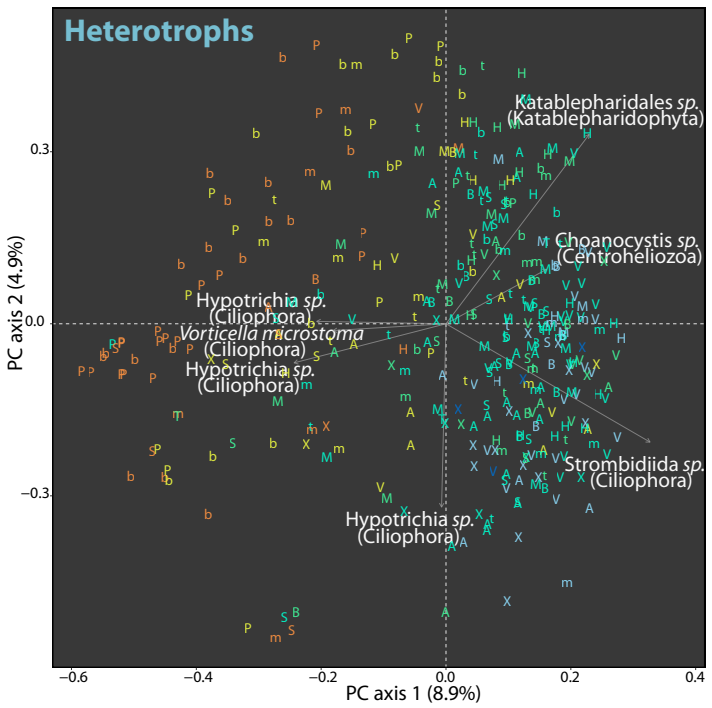


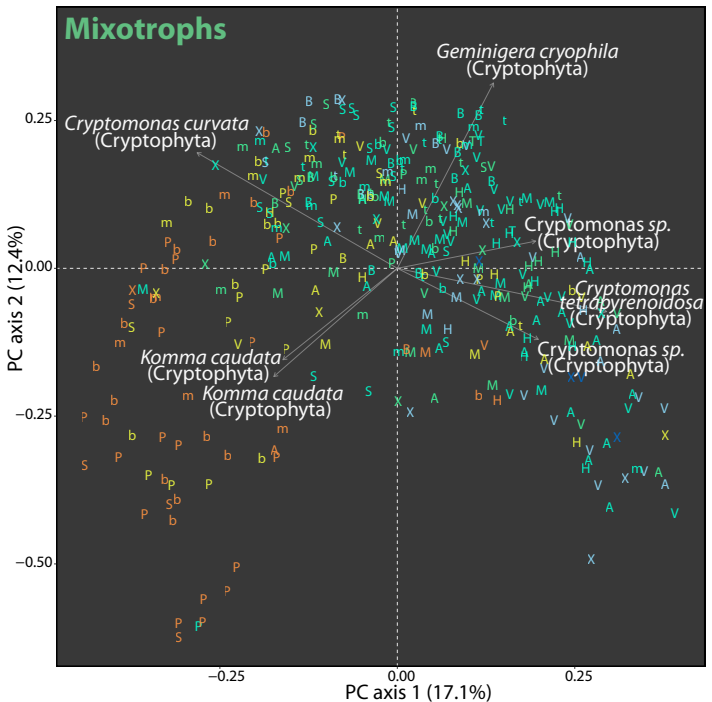












Ecozone

- T Taiga Cordillera
- B Boreal Cordillera
- m Montane Cordillera
- X Pacific Maritime
- t Taiga Plains
- S Semi–Arid Plateaux
- b Boreal Plains
- P Prairies
- M Mixedwood Plains
- V Boreal Shield
- H Atlantic Highlands
- A Atlantic Maritime

Trophic state

- Hypereutrophic
- Eutrophic
- Mesoeutrophic
- Mesotrophic
- Oligotrophic
- Ultraoligotrophic