LIMNOLOGY and OCEANOGRAPHY

ASLO

Limnol. Oceanogr. 9999, 2022, 1–16

doi: 10.1002/lno.12247

Deterministic processes drive national-scale patterns in lake surface sediment bacteria and eukaryotic assemblage composition

John K. Pearman [1]* Janet Adamson [1] Georgia Thomson-Laing [1] Lucy Thompson [1] Sean Waters [1] Marcus J. Vandergoes [2] Jamie D. Howarth [3] Susanna A. Wood [1]

[1]Coastal and Freshwater Group, Cawthron Institute, Nelson, New Zealand
[2]GNS Science, Lower Hutt, New Zealand
[3]University of Victoria, Wellington, New Zealand

Abstract

Biological communities within lake surface sediments play a vital role in biogeochemical cycling and ecosystem services. Knowledge on abundance-occupancy patterns and assembly processes across large spatial scales and over multiple environmental gradients is limited, yet essential to aid in protection and restoration. In the present study, surface sediment samples were collected from 296 lakes across a wide spatial scale and covering multiple interacting environmental gradients including size, depth, altitude, and trophic state. A suite of physicochemical parameters were used to characterize the environmental conditions and bacterial and eukaryotic assemblages were determined using *16S* and *18S rRNA* metabarcoding. The majority (~ 55%) of amplicon sequence variants were only found in a single lake with eukaryotes having a more restricted distribution than bacteria. Deterministic processes were inferred to be dominant for both bacteria (78%) and eukaryotes (51%), with variable selection being especially important for bacteria (49%). Variation partitioning indicated that land use in the catchment, which is strongly related to trophic state, was the most important environmental factor in explaining the assemblage composition. This nationwide study across broad gradients provides new insights into the ecology of bacteria and eukaryotes in lake surface sediments and a platform to better understand the effects of multiple environmental stressors on lake sediment assemblages.

Lakes accumulate nutrients and environmental contaminants from the surrounding landscape and atmosphere. This makes them especially vulnerable to anthropogenic and natural perturbations and because of this they are often considered sentinels of environmental change (Adrian et al. 2009). Humans have used lakes and their catchments as a resource for millennia, exposing them to multiple stressors including habitat degradation, pollution, and introducing non-native species (Dudgeon et al. 2006). These stressors have had profound effects on lake ecosystems leading to global declines in their health (McCrackin et al. 2017). Understanding how the biological communities of lakes are affected by environmental and anthropogenic stressors and other factors is critical to ensure appropriate protection and enable restoration where required.

Lakes also provide a unique opportunity to study biogeographical patterns and their drivers because they are discrete bodies of water, which are inherently connected to their surrounding terrestrial habitats. However, even within reasonably short geographic distances they can experience diverse environmental conditions due to differences in parameters such as size, depth, geomorphology, and land use and geology in their catchment. The plankton communities of lakes have been extensively studied (Kraemer et al. 2020; Schallenberg et al. 2021; Pearman et al. 2022a). Work involving sediments, has generally focused on paleolimnological studies, involving taxa that fossilize such as diatoms (Rühland et al. 2003) and chironomids (Irvine et al. 2012), although molecular approaches to assess other taxa are increasing (Capo et al. 2022).

Specific studies on communities in surface sediments have received far less attention using molecular methods, despite their importance in many biogeochemical processes (Forsberg 1989). To date most studies of sediment communities have focused on one or a few lakes within a relatively

*Correspondence: john.pearman@cawthron.org.nz

Additional Supporting Information may be found in the online version of this article.

Author Contribution Statement: J.K.P., M.J.V., J.D.H., and S.A.W. designed the research with all authors taking part in the sample collection. J.A., G. T.-L., S.W., and L.T. undertook the laboratory work while J.K.P. and S.A. W. analyzed the data and wrote the manuscript with contributions from the other authors.

confined geographic region (Zeng et al. 2019; Jiao et al. 2020; Pearman et al. 2020).

Macroecological-based studies have shown that most rare species are narrowly distributed whereas abundant species tend to be widespread (Gaston et al. 2000). This abundance–occupancy concept has now been applied to microbial systems where similar patterns have been observed (Shade et al. 2018). Research indicates that deviations away from these trends are the result of environmental variables driving changes in community structure (Shade et al. 2018), for example, when conditions promote phytoplankton blooms (Shade et al. 2014).

In general, the expectation is that the structure of biological assemblages will be determined by a balance of deterministic and stochastic processes (Zhou and Ning 2017; Liu et al. 2019; Sadeghi et al. 2021). Deterministic processes are based on the principal that ecological selection acts on fitness differences among taxa (Vellend 2010). Deterministic processes can be split into two classes. If environmental conditions vary spatially and/or temporally, high variation in the community structure can exist, referred to as variable selection (Zhou and Ning 2017). In contrast, if environmental conditions are homogeneous little variation in community structure is expected, this is known as homogeneous selection (Zhou and Ning 2017). Abiotic factors impacting lakes include within lake factors (e.g., pH, nutrient concentrations, temperature; Ruuskanen et al. 2018; Pearman et al. 2020) and effects derived from the wider catchment such as land use (Kraemer et al. 2020) and altitude (Li et al. 2017; Pearman et al. 2020). Biotic interactions (e.g., predation and competition) can also have deterministic effects on biological communities (Zhou and Ning 2017). Stochastic processes can become predominant in systems where taxa are functionally equivalent and not strongly affected by environmental variables with dispersal and ecological drift dominating assembly processes (Sloan et al. 2006). Stochastic processes can be further split into dispersal limitation and homogenizing dispersal. Limitations in dispersal coupled with genetic drift can increase community variation (Stegen et al. 2015). In contrast, high levels of dispersal result in a more homogenous community and are referred to as homogenizing dispersal (Zhou and Ning 2017).

Over the last two decades, high-throughput sequencing technologies have enabled in-depth investigations into the composition of biological assemblages, facilitating new insights into their biodiversity, distribution, and structure (Compson et al. 2020). Most studies on lakes have focused on the plankton including zooplankton (Duggan 2007), eukaryotic phytoplankton (Soininen et al. 2011; Logares et al. 2018), and planktonic bacteria (Kraemer et al. 2020; Li et al. 2020; Schallenberg et al. 2021). Because planktonic communities are susceptible to rapid fluctuations in environmental conditions (e.g., weather events, Ji et al. 2019) and physical processes such as lake mixing, regular sampling is required to obtain insights into assemblage composition over extended, that is, annual, time periods. In contrast, lake surface sediments incorporate living organisms who inhabit the sediment as well as organisms that have sedimented from the water column providing an integrated record of past conditions within a lake. Studies on bacterioplankton have shown that deterministic processes and especially homogeneous selection is prevalent (Zeng et al. 2019; Jiao et al. 2020) although stochastic processes were shown to be important in a large-scale investigation of bacterioplankton in Canada (Kraemer et al. 2020). The investigation of assembly processes within surface sediments has received less attention and only been undertaken on small scales; for example, Zeng et al. (2019) showed that variable selection was prevalent for bacteria across 10 Chinese lakes and Jiao et al. (2020) showed a similar response for 13 shallow lakes in China. Knowledge on the relative importance of these processes, not just on bacteria but also the eukaryotic component of surface sediments, across a large number of lakes covering a wide spatial area is lacking and is vital in ensuring appropriate mitigation actions are taken to protect or restore these habitats.

To advance knowledge on the distribution patterns of bacteria and eukaryotes within surface sediments of lakes and the processes driving these patterns, we implemented a carefully designed sampling strategy and sampled surface sediments from 296 lakes throughout New Zealand spanning 12° of latitude. The selected lakes covered a range of environmental gradients including size, depth, altitude, trophic status, and catchment land use, in order to establish the contribution of deterministic and stochastic processes on structuring the bacterial and eukaryotic assemblages. We set out to test three hypotheses: (i) bacteria and eukaryotes would have a positive abundancy-occupancy pattern but this would be more pronounced among bacteria due to the greater number of amplicon sequence variants (ASVs) detected; (ii) deterministic processes would be inferred to be the dominant processes in structuring bacterial and eukaryotic sediment assemblages with variable selection more important than homogeneous selection; and (iii) differing environmental variables would drive the distribution of the abundant bacteria and eukaryotic taxa, and this will largely be attributed to different physiological requirements between these taxa.

Methods

Study lakes

A total of 296 lakes were sampled between October 2018 and April 2021 (Fig. 1; Supporting Information Table S1). Lakes were selected to ensure they were representative of the wide range of gradients observed across all New Zealand's 3821 lakes (Supporting Information Fig. S1). Lakes ranged in altitude from sea level (e.g., Lakes Moawhitu and Kohangapiripiri) to 1839 m (Duncan Stream Tarn) above sea level, and in area from 1 ha (Lake Kawau and Dukes Tarn) to 29,825 ha (Lake Wakatipu). The depths of the lakes ranged

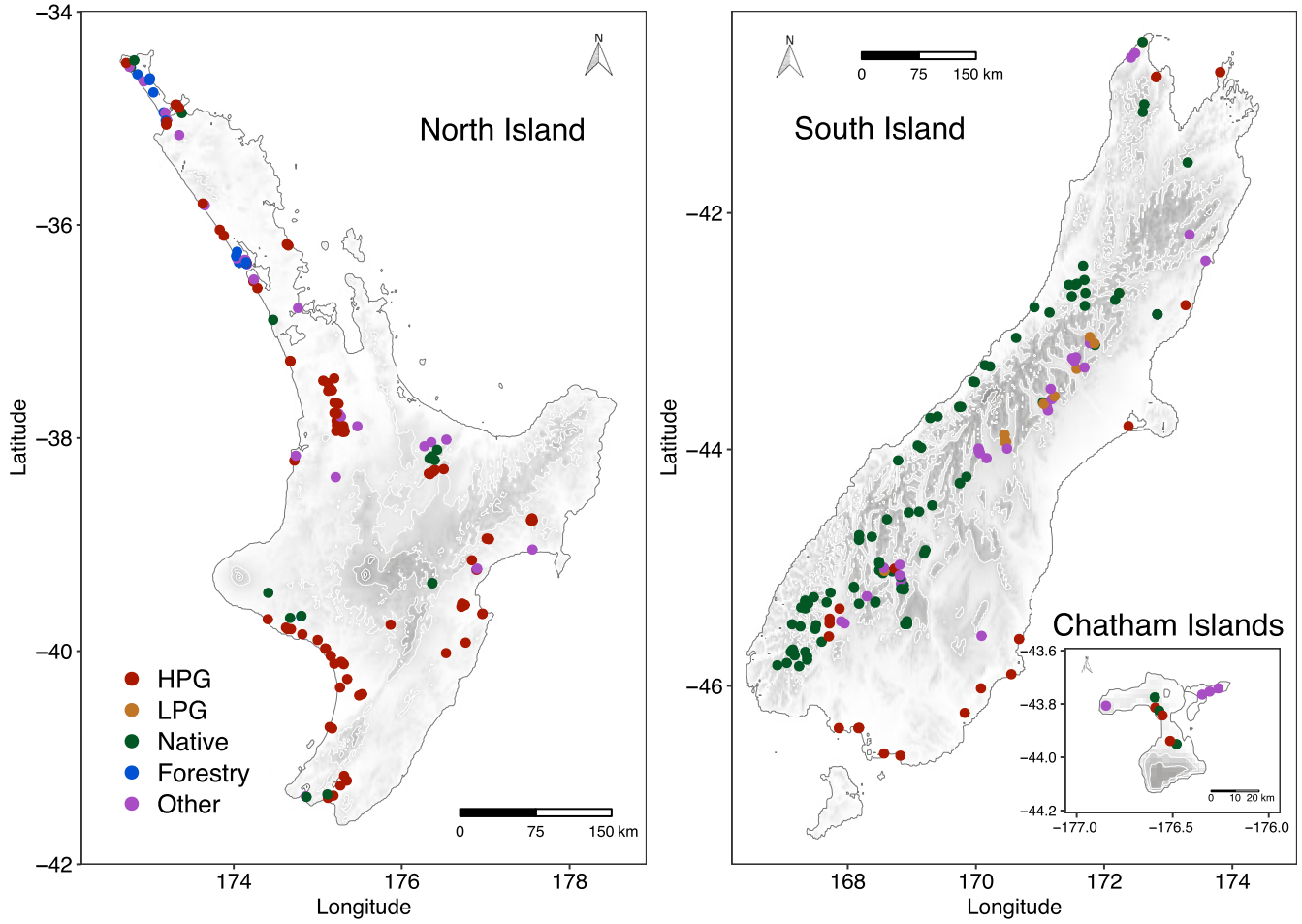

Fig. 1. Location of the 296 lakes sampled, including those on offshore Chatham Island. Lake points are colored by the dominant (> 50%) land use in the catchment. HPG, high-productivity grassland, LPG, low-productivity grassland. Other includes lakes with a catchment dominated by another land use category or one where there is no land use > 50%.

from 0.2 m (Sutton Salt Lake) to 444 m (Lake Manapouri; Supporting Information Table S1). Prior to sampling, the deepest part of the lake was identified using a side scan sonar survey (Lowrance) or depth sounder (HawkEye H22PX). Sampling was undertaken at this position except for lakes deeper than 100 m where sampling was undertaken in shallower bays (Supporting Information Table S1).

Lake, catchment, and land-use data

Lake depth and Secchi disk depth were assessed in situ. Catchment descriptors were extracted from the Freshwater Ecosystems of New Zealand database (Leathwick et al. 2010). Eight land cover variables were derived from the most recent satellite imagery available in the Land Cover Database Version 5 (Landcare Research New Zealand Ltd, https://lris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-zealand/): (1) native vegetation, (2) urban, (3) non-native vegetation, (4) water, (5) forestry, (6) high production grassland, (7) low producing grassland, and (8) other. The specific land use characteristics that make up these eight broad groups are described in Supporting Information Table S2.

Sample collection, nutrient, and elemental characterization of surface sediment

At each site, ponar grabs were used to collect triplicate surface sediment samples. Surface sediment samples (top 5 mm, ca. 1 g) were taken using sterile spatulas. These were placed in LifeGuard™ Soil Preservation Solution (3 mL, Qiagen) and stored frozen (−80°C) for later DNA extraction. The remaining top 2 cm from the three ponar grabs was collected using spatulas and placed in 500-g containers. This was homogenized, stored chilled (4°C), and shipped to the laboratory within 48 h for nutrient and elemental characterization.

Samples (1 liter) were collected at the surface for water column chlorophyll *a* (Chl *a*) at the same point in the lake as the ponar grabs and kept on ice until further processing. Filtration (up to 600 mL) was undertaken using GF/C (Whatman) filters.

Filters were placed in aluminum foil and stored in the dark (−20°C). Chl *a* analysis was undertaken at Watercare Laboratories (Auckland, New Zealand) following the APHA 10200 H method with a reporting limit of 0.0006 mg L^{-1}.

Water chemistry samples (1 L) were collected in the surface mixed layer, as determined by a RBRmaestro Multi-Channel Logger (RBR), of the lake using an integrated sampler. Subsamples (40 mL) were taken for total nitrogen (TN), total phosphorus (TP), and total organic carbon (TOC) with 40 mL for each analysis. Dissolved organic carbon (DOC; Supporting Information Table S1) samples were obtained by filtering 40 mL through a low bleed 0.45-μm filter. The APHA 4500 method was used to measure total nutrients on a flow injection analyzer (reporting limits in Supporting Information Table S2). TOC and DOC analyses were undertaken by combustion analysis at 850°C using APHA 5310 B methods (Supporting Information Table S2).

For sediment nutrient and elemental characterization porewater was decanted from the sediment after centrifugation (3000 × *g*, 40 min, 4°C). The sediment metals iron (Fe), manganese (Mn), aluminum (Al), calcium (Ca), lead (Pb), copper (Cu), zinc (Zn), cadmium (Cd), phosphorus (P), and sulfur (S) were analyzed using Inductively Coupled Plasma-Mass Spectrometry analysis based on the US Environmental Protection Agency (EPA) method 200.8. Reporting limits are detailed in Supporting Information Table S2. Catalytic combustion at (900°C, O_2) and separation using a thermal conductivity detector was used to measure sediment TN and TOC. These data were only available for 190 samples (Supporting Information Table S1).

DNA extraction, polymerase chain reaction, high-throughput sequencing, and bioinformatics

All molecular analyses (i.e., DNA extraction, polymerase chain reaction [PCR] set-up, template addition, PCR analysis), were conducted in separate UV sterilized dedicated laboratories to minimize potential cross-contamination. Laminar flow cabinets with HEPA filtration were used for PCR set-up and template addition.

DNA was extracted from surface sediment subsamples (0.25 g wet weight of 1 g sample of the top 5 mm) using the DNeasy PowerSoil Kit (Qiagen) following the manufacturer's instructions using a QIAcube extraction robot (Qiagen). A negative extraction control (nuclease-free water) was included every 23rd sample.

PCR was used to amplify the V3–V4 regions of the bacterial *16S rRNA* gene using the 341F and 805R primers (Herlemann et al. 2007; Klindworth et al. 2013) and the V4 region of the eukaryote nuclear *18S rRNA* gene using the Uni18SF and Uni18SR primers (Zhan et al. 2013). PCR conditions and library construction are as described in Pearman et al. (2020). Samples were sequenced on an Illumina Miseq™ platform at the Auckland Genomics Facility. Raw sequence reads are deposited in the NCBI short read archive under the accession number: PRJNA606991, PRJNA750120, and PRJNA813318. A small geographically restricted subset of the raw data was previously analyzed in Pearman et al. (2020), while the bacterial ASV data from the majority of the lakes were used for the development of a sediment bacterial trophic index (Pearman et al. 2022*b*).

Primers were trimmed from the raw reads using cutadapt with a single mismatch allowed (Martin 2011). ASVs were inferred using the DADA2 package (Callahan et al. 2016) within R (R Core Team 2020). A maximum number of "expected errors" (maxEE) threshold of two (forward reads) and four (reverse reads) was used with reads truncated to 230 and 228 base pairs (bp) for forward and reverse reads, respectively. Sequence variants were determined for the forward and reverse reads based on a parametric error matrix constructed from the first 10^8 bp. Pair-end reads, after the removal of singletons, were merged with a maximum mismatch of 1 bp and a required minimum overlap of 10 bp. Chimeric sequences were removed using the script removeBimeraDenovo within the DADA2 package.

Taxonomic classification of the ASVs was undertaken against the SILVA 138 (Pruesse et al. 2007) and PR2 databases (Guillou et al. 2013) for the *16S* and *18S rRNA* gene datasets, respectively. The RDP classifier (Wang et al. 2007) was used with a bootstrap of 70 to enable classifications at higher taxonomic levels.

The results were combined into a *phyloseq* object (McMurdie and Holmes 2013) and ASVs not assigned at kingdom level removed. For the *16S rRNA* dataset sequences assigned as eukaryotes, chloroplasts and mitochondria were removed and bacteria and the phylum Chordata were removed from the *18S rRNA* dataset. To assess potential contamination, negative controls were evaluated. Each sequencing run was processed separately with the number of reads observed in negative controls of the sequencing run removed via subtraction from the corresponding samples. ASVs that were present in only a single replicate within a lake were considered as potentially erroneous and removed from that lake's data.

For comparisons between lake samples, triplicates were combined together and subsampling to an even depth was undertaken for each sample at 12,579 and 18,459 reads for bacterial and eukaryotic datasets, respectively. Samples which did not reach this threshold were removed leaving 285 bacterial and 281 eukaryotic samples, of which 270 were shared between the two datasets. For some analysis a microbial subset of the eukaryotes was used. This was assembled by selecting microbial groups based on taxonomy from the rarefied eukaryotic dataset.

Statistical analysis

Principal component analysis was undertaken on the environmental data within the base package of R to give an indication of the environmental gradients sampled. The results were

visualized with *ggfortify* v0.4.14 (Horikoshi et al. 2022) and *ggplot2* v3.3.6 (Wickham 2016).

The taxonomic structure and composition of overall lake assemblages were assessed by calculating both the number of ASVs per class as well as the relative proportions and visualized in *ggplot2* v3.3.6 (Wickham 2016). Occupancy of ASVs was calculated using R, based on the number of lakes (triplicate samples were bioinformatically combined) within which an ASV occurred.

To infer the relative contributions of deterministic and stochastic processes in structuring bacteria, microbial eukaryotic and eukaryotic, sediment assemblages the ecological modeling frameworks developed and adapted by Stegen and colleagues were employed (Stegen et al. 2013, 2015). Initially, the assemblage was tested for phylogenetic signal against environmental variables using mantel correlog in *vegan* and as there were significant signals we proceeded with the analysis. Pairwise phylogenetic turnover between assemblages was assessed by calculating the mean-nearest-taxon-distance (ßMNTD; Fine and Kembel 2011). This was achieved by constructing a phylogenetic tree, using the package *msa* v1.24.0 (Bodenhofer et al. 2015) and *phangorn v2.8.1* (Schliep 2011), incorporating "abundant" ASVs (those with at least 10 reads across the dataset). The phylogenetic distance between each ASV in an assemblage and its nearest relative in a second assemblage was then quantified. Under the assumption that ecological selection was not the primary cause of differences in pairs of assemblages a null distribution of ßMNTD was calculated using the R packages *picante v1.8.2* (Kembel et al. 2010) and *iCAMP* v1.3.4 (repetitions = 999, Ning et al. 2020). By comparing the null model with the observed ßMNTD values and normalizing by the standard deviation, the beta-nearest-taxon-index (ßMNTI) was calculated. Deterministic processes are indicated by deviations away from null distributions. If environmental conditions are similar, selective pressures will be consistent and homogenous selection will be dominant, resulting in low levels of change in the assemblage and ßMNTI values < -2. Conversely, variable selection is indicated by a ßMNTI > 2 and is suggestive of among taxa fitness differences to environmental changes resulting in higher than expected pairwise differences in assemblages. Pairwise comparisons which did not deviate from the null distribution indicated that deterministic processes were weak and thus stochastic processes, dispersal limitation and homogenizing dispersal, were evaluated. Stochastic processes were assessed by calculating the Raup–Crick metric adapted to account for species relative abundances (RC_{Bray}). A null model was calculated, and the observed values compared against it and standardized to between -1 and 1 (Stegen et al. 2015). If deterministic processes are low, then large differences are inferred to be primarily due to dispersal limitation and subsequent compositional drift of assemblages and is indicated by $RC_{Bray} > 0.95$. High rates of dispersal, called homogenizing dispersal, can be inferred to be the primary cause of similarity between communities if deterministic processes are at a low level and are denoted by RC_{Bray} values < -0.95. Values between -0.95 and 0.95 are interpreted as having no dominant assembly process.

A total of 31 environmental variables (catchment characteristics and physiochemical variables) were obtained (Supporting Information Table S1). These variables were assessed for co-linearity with selection of one of the co-linear variables occurring based on ecological knowledge (e.g., water column TOC and DOC were co-linear, therefore, DOC was selected for analysis as this compound is likely to be bioavailable to microorganisms; Supporting Information Table S2). Remaining environmental variables were scaled and centered, using the R package *caret v6.0* (Kuhn 2008). Any missing values for the environmental variables were imputed via bagging using the caret package. This method fits a bagged tree model for each predictor (as a function of all others) and imputes a value for the missing variable based on this model. A total of 13 environmental variables were retained for analysis (Supporting Information Table S2). As appropriate water quality data were not available for the majority of lakes, trophic state was determined for each lake based on the surface sediment bacterial trophic index (SBTI) as detailed in Pearman et al. (2022*b*). This index is based on the proportion of bacterial trophic state indicator ASVs in the lake. The indicator ASVs were determined based on the investigation of the bacterial assemblage in 96 monitored lakes which had a trophic lake index assigned to them (Pearman et al. 2022*b*). The SBTI was calculated by the following equation:

$$\begin{aligned}\mathrm{SBTI} = {} & 0.017 \times \%\mathrm{microtrophic} + 0.026 \times \%\mathrm{oligotrophic} + \\ & 0.031 \times \%\mathrm{mesotrophic} + 0.040 \times \%\mathrm{eutrophic} + \\ & 0.055 \times \%\mathrm{supertrophic} + 0.067 \times \%\mathrm{hypertrophic},\end{aligned}$$

where, for example, %microtrophic is the % of indicator taxa in the sample that are microtrophic indicators. The SBTI value was then classified into six trophic states: microtrophic (SBTI 1–2), oligotrophic (SBTI 2–3), mesotrophic (SBTI 3–4), eutrophic (SBTI 4–5), supertrophic (SBTI 5–6), and hypertrophic (SBTI > 6). The microtrophic category would fall into the ultra-oligotrophic of the trophic state index (TSI) (Carlson 1977), while both supertropic and hypertrophic would be classified as hypereutrophic in the TSI.

Moran's eigenvector maps (MEMs) were created and used as spatial variables. The distance between two lakes were calculated with the R package *geosphere* v1.5 (Hijmans et al. 2017), and these distance were used as the input for the calculation of the MEMs using the function dbmem in the package adespatial v0.3 (Dray et al. 2016). Environmental variables were scaled and centered, and MEMs were combined and important variables were selected based on distance-based redundancy analysis (dbRDA; McArdle and Anderson 2001) stepwise selection (forward) using the function ordiR2step in the package *vegan v2.5.7* (Oksanen et al. 2007).

Variation partitioning was undertaken using the environmental variables and MEMs that were selected during the stepwise selection. The environmental variables were split into three categories (lake/water column variables, sediment variables, and catchment variables, shown in Supporting Information Table S2). The variation partitioning was undertaken using the Bray–Curtis distance matrices as the response variable using the function varpart in the package *vegan* (Oksanen et al. 2007).

Mantel tests (method = Spearman; permutations = 999) were used to assess whether the bacterial and eukaryotic assemblages showed similar patterns based on Bray–Curtis dissimilarities.

To investigate the relationship between bacteria and eukaryotic assemblages amalgamated at the genera level, a dbRDA based on Bray–Curtis dissimilarity was undertaken using the scaled and centered environmental characteristics and MEMS. Stepwise selection (forward) was undertaken to select the important variables.

Correlation analysis was undertaken using data collated at the genera level. Genera were considered abundant if they had a relative abundance of greater than 1.5% in at least 10% of lakes. We selected this threshold to allow the key patterns in abundant genera to be observed, noting that over 200 genera were present in over 10% of lakes. Correlations among genera were considered significant if the adjusted *p*-value was less than 0.05. Correlations between abundant genera and environmental variables were calculated using rcorr in *Hmisc v4.6-0* (Harrell and Dupont 2021) package using Spearman's rank using the raw untransformed environmental data. Lakes with missing data for a particular environmental variable were not considered. Correlations were considered significant if the adjusted p values (method = Benjamini Hochberg) were less than 0.05.

Results

Environmental parameters

The environmental data showed a distinct split in the characteristics of the lakes among islands. The South Island has predominantly higher altitude lakes that were of glacial origin with native vegetation in their catchments and a higher Secchi disk depth. In contrast the lakes of the North Island generally have a higher nutrient and organic carbon content with higher amounts of high-productivity grassland in their catchment (Fig. 2).

General diversity patterns

After rarefaction, there were 73,910 bacterial and 10,737 eukaryotic ASVs. On average, there were 1090 (range 263–2438) bacterial and 144 (15–398) eukaryote ASVs per lake.

Gammaproteobacteria was the dominant bacterial class accounting for 17.1% of ASVs and 21.8% of reads followed by Bacteroidia (13.6% ASVs and 13.2% reads; Fig. 3). Cyanobacteria while only being 25th in rank in terms of number of ASVs was the 5th highest class in terms of proportion of reads accounting for 3.7%. Phototrophs dominated the eukaryote assemblage and were predominant in terms of both proportion of ASVs and reads with Dinophyceae (19.2% and 29.7%) and Chlorophyceae (10.6% and 6.7%) most abundant (Fig. 3). Clitellata and Ostracoda had relatively low proportions of ASVs (1.1% and 0.6%, respectively) but accounted for a substantial component of the assemblage in terms of proportion of reads (8.0% and 6.5%).

Occupancy

For both bacteria and eukaryotes, the majority of ASVs (bacteria: 39,967 [54%] and eukaryotes: 5866 [55%]) were found in a single lake. Only 41 (0.06%) bacterial ASVs and 2 (0.02%) eukaryotic ASVs were present in greater than 50% of lakes. The highest lake occupancy for an ASV (i.e., the ASV was found in this percentage of lakes) was 84% for bacteria, belonging to the genus Sva0081 (family Desulfosarcinaceae) and 56% for eukaryotes, belonging to the class Cryptomycota. In general, the mean abundance of ASVs was higher for ASVs with a greater occupancy for both the bacterial and eukaryotic assemblages (Fig. 4).

Deterministic verses stochastic processes

Deterministic processes (variable and homogeneous selection) were inferred to be most prevalent for both bacteria and eukaryotes accounting for 78% and 50% of lake pairwise comparisons, respectively (Table 1). For bacteria variable selection dominated (49%), while for eukaryotes variable selection (25%) and homogeneous selection (25%) had a similar contribution to assembly overall. When only analyzing the microbial eukaryotes a slightly different pattern was observed to the eukaryotes overall, with a higher inferred contribution from homogeneous selection than variable selection (Table 1). Eighty-five percent of the lake pairwise comparisons that were inferred to be dominated by variable selection for the bacterial assemblage were between lakes with different trophic state. For those dominated by homogeneous selection 30% of the lakes had a similar trophic state. A similar pattern was observed for the eukaryotes with 85.7% of the variable selection pairwise comparisons having a different trophic state and 73.5% for those dominated by homogeneous selection. Stochastic processes were only inferred as the dominant assembly process in less than 10% of lake pairwise comparisons for bacteria and eukaryotes. For the eukaryotic assemblages, a high proportion (40%) of the pairwise comparisons had no dominant process explaining assembly. A similar pattern was also observed for the microbial eukaryote component.

Variation partitioning of the bacterial assemblage showed that while the spatial component individually accounted for 3.2% of the variation, the combined environmental components (when not shared with the spatial component) accounted for 5.3%, with the catchment land use

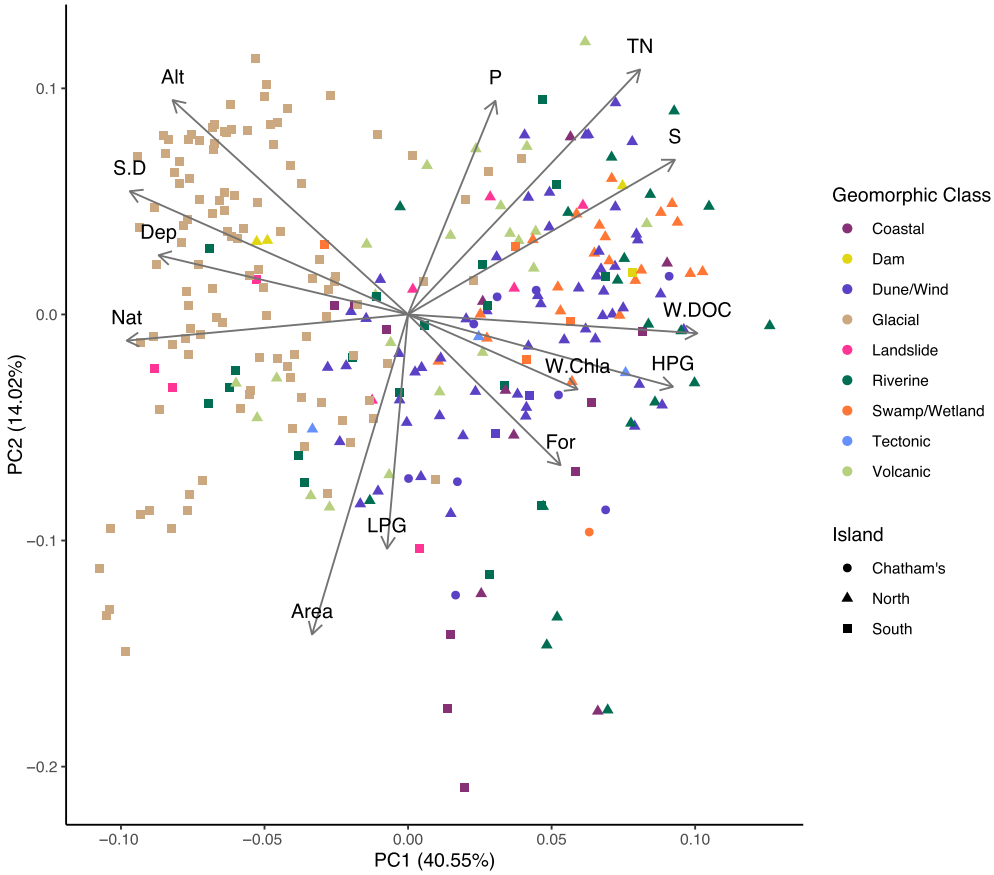

Fig. 2. Principal component analysis (PCA) of the environmental data. Points are colored by the geomorphic class and the shape of the point indicates which island the lake is found on. Alt, altitude; Area, lake area; Dep, maximum lake depth; For, forestry in the catchment; HPG, high-productivity grassland in catchment; LPG, low-productivity grassland in catchment; Nat, native vegetation in catchment; S.D., Secchi disk depth; W.Chla, water column Chl *a*; W.DOC, water column dissolved organic carbon.

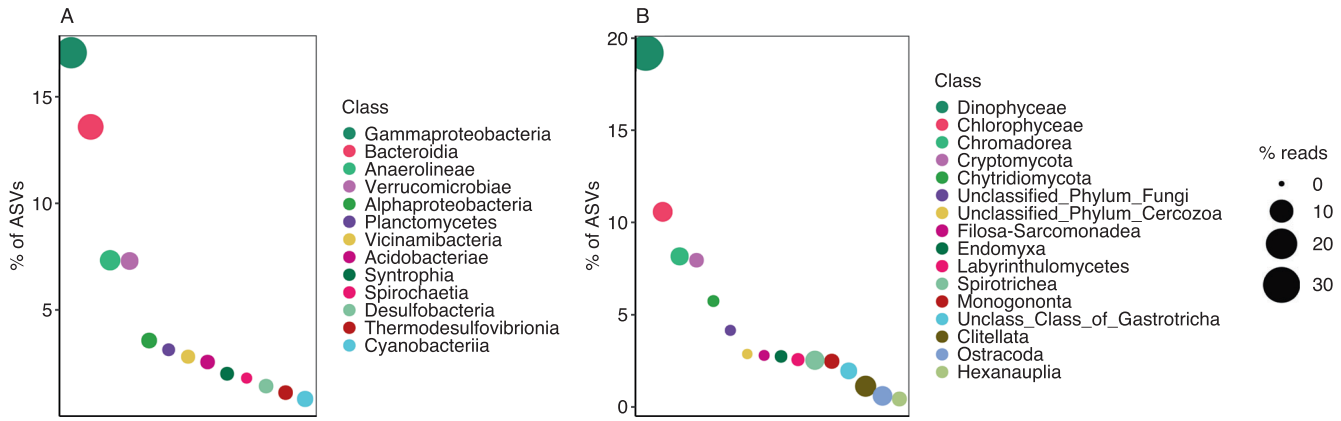

Fig. 3. Percentage of ASVs and reads per taxonomic class for; (**A**) bacteria and (**B**) eukaryotes. The order of taxa along the *x*-axis is based on the percentage of ASVs attributed to each class. Only the taxa that were ranked in the top 10 for percentage of ASVs or abundance are depicted.

characteristics accounting for the highest proportion (2.1%) of the variation explained by the environmental data. Most of the variation (87.7%) could not be attributed to the investigated variables. For the eukaryotic assemblage, a similar pattern was evident with the majority of the variation (89.4%) not attributed. Catchment land use (2.6%) and spatial factors

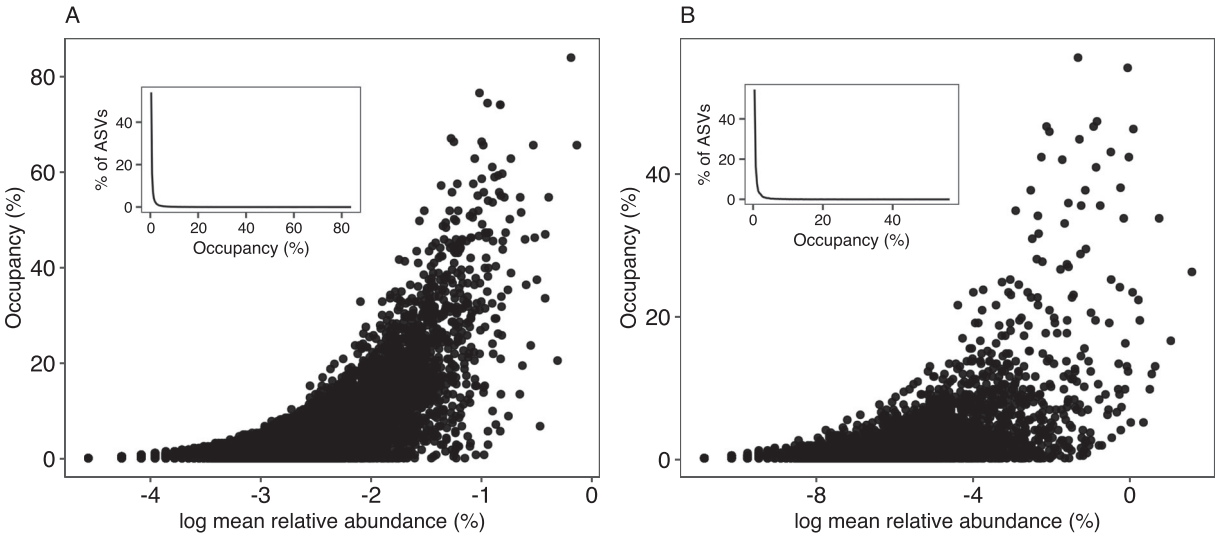

Fig. 4. Abundance-occupancy relationships for (**A**) bacteria and (**B**) eukaryotes. Inserts in the figures show the percentage of ASVs observed in the proportion of lakes surveyed.

Table 1. The proportion of lake pairwise comparisons attributed to each assembly process contributing to the composition of bacterial, eukaryotic and microbial eukaryotes assemblages. Assembly processes are based on those of Stegen et al. (2013, 2015). Only ASVs with greater than or equal to 10 reads were used for the analysis.

	Deterministic		Stochastic		
	Variable selection (%)	Homogenous selection (%)	Dispersal limitation (%)	Homogenizing dispersal (%)	No dominant process (%)
Bacteria	49	29	6	2	14
Eukaryotes	25	25	9	1	40
Microbial eukaryotes	15	37	7	1	41

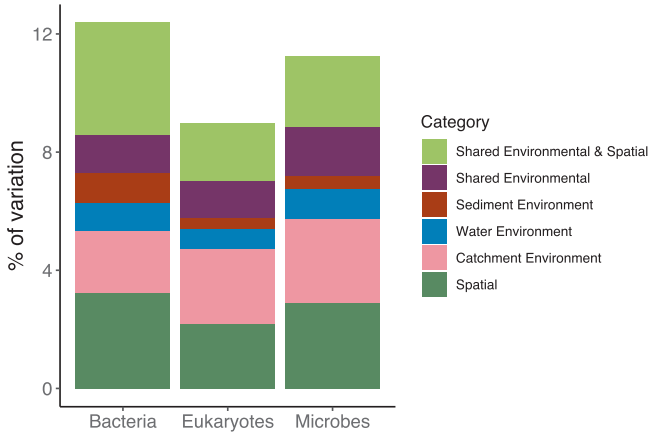

Fig. 5. The contribution of different factors to explaining the variation in the biological assemblages. Microbes refers to eukaryotic microbes.

(2.2%) accounted for similar amounts of variation individually. Overall environmental factors (without spatial impacts) contributed 5.1% to explaining the variation (Fig. 5).

Mantel tests indicated that there was a significant positive relationship between the bacterial and eukaryotic assemblages (Mantel $r = 0.5417$; $p < 0.001$) based on Bray–Curtis dissimilarities.

Environmental drivers

dbRDA at the genus level showed that measured variables only explained a small percentage of the variation for both bacteria (14.5% first two axes; Fig. 6A,B) and eukaryotes (9.5% first two axes; Fig. 6C,D). Overall, the dbRDA was significant for both bacteria ($F = 4.36$; $p < 0.001$) and eukaryotes ($F = 3.32$; $p < 0.001$) with all terms for bacteria also significant. For the eukaryotes all terms except for the spatial terms MEM35 ($p = 0.055$) and MEM16 ($p = 0.069$).

To assess the environmental drivers for specific abundant genera, Spearman's Rank correlation analysis was undertaken. *Nitrospira* was associated with lakes of lower trophic state (better water quality) and negatively correlated to TN, water column Chl *a* and DOC as well as latitude and the proportion of high-productivity grassland in the catchment (Figs. 6A,B, 7A).

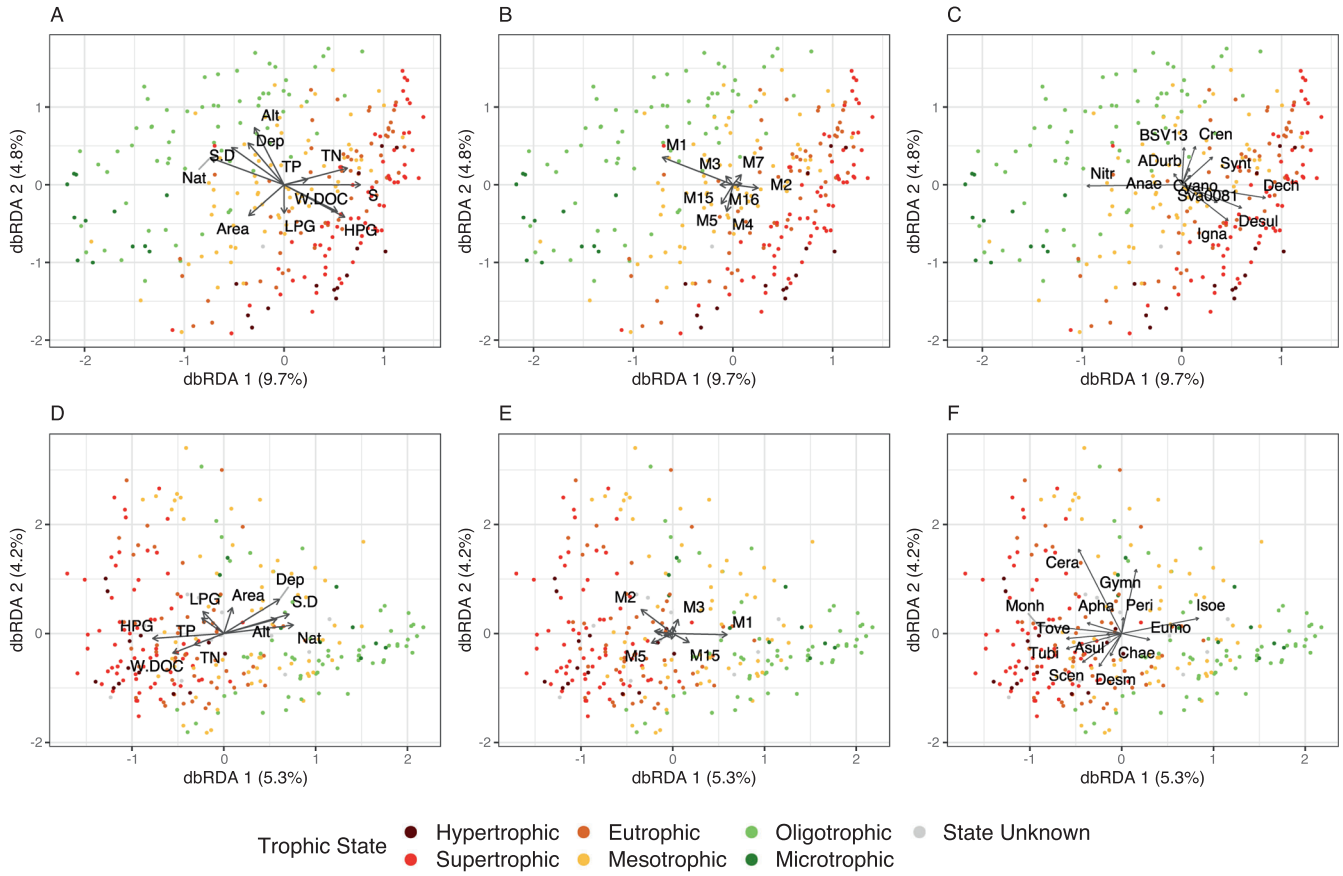

Fig. 6. dbRDA based on Bray–Curtis dissimilarity matrix on the sediment assemblage at the genus level showing environmental, spatial and abundant taxa relationships for bacteria (**A**–**C**), and eukaryotes (**D**–**F**). Points are colored by predicted trophic state calculated based on the bacteria assemblage based on Pearman et al. (2022*b*). Alt, altitude; Area, lake area; Dep, maximum lake depth; HPG, high-productivity grassland in catchment; Lat, latitude; LPG, low-productivity grassland in catchment; Nat, native vegetation in catchment; S.D., Secchi disk depth; W.Chla, water column Chl *a*, W.DOC, water column dissolved organic carbon. Only significant environmental variables shown. The spatial vectors are represented by Moran eigenvector maps. A total of 15 spatial variables were selected by forward selection during the dbRDA analysis with the spatial autocorrelation depicted geographically in Supporting Information Fig. S2. ADurb, ADurb.Bin063-1; Apha, *Aphamonas*; Anae, *Anaeromyxobacter*; Asul, *Asulcocephalium*; Cera, *Ceratium*; Chae, *Chaetonotus*; Cren, *Crenothrix*; Cyano, *Cyanobium* PCC-6307; Dech, *Dechloromonas*; Desm, *Desmodesmus*; Desul, Desulfatiglans; Eumo, *Eumonhystera*; Gymn, *Gymnodinium*; Igna, *Ignavibacterium*; Isoe, *Isoetes*; Monh, *Monhystera*; Nitr, *Nitrospira*; Peri, *Peridinium*; Scen, *Scenedesmus*; Synt, *Syntrophus*; Tove, *Tovellia*; Tubi, *Tubificoides*.

In contrast *Dechloromonas*, *Desulfatiglans*, *Ignvibacterium*, and Sva0081 were more prevalent in lakes with higher trophic states (reduced water quality) associated with higher concentrations of DOC and Chl *a* in the water column, increased sediment TN and TP, and greater proportions of modified catchments (high-productivity grassland, low-productivity grassland, and forestry; Figs. 6A,B, 7A; Supporting Information Table S3). The genus *Cyanobium* PCC-6307 was positively associated with sediment TN and low-productivity grassland in the catchment and negatively correlated to lake depth, area, and latitude (Figs. 6A,B, 7A; Supporting Information Table S3).

For eukaryotes, the relative abundance of the macrophyte genus *Isoetes* was positively correlated with Secchi disk depth, lake area and depth, the proportion of native vegetation and low-productivity grassland in the catchment and altitude. *Isoetes* was negatively correlated with latitude as well as TN and TS in sediment, water column Chl *a* and DOC, and high-productivity grassland in the catchment (Figs. 6C,D, 7B; Supporting Information Table S4). Chlorophyta genera *Scenedesmus* and *Desmodesmus* were both negatively correlated with Secchi disk depth, maximum lake depth, lake area, and native vegetation in the catchment and positively correlated to sediment TN and water column DOC and Chl *a* (Figs. 6C,D, 7B; Supporting Information Table S4). Dinoflagellata genera showed differing correlation patterns. *Tovellia* and *Asulcocephalium* were more abundant in lakes of a higher trophic state with higher nutrients and modified catchments (Fig. 5B). In contrast, *Gymnodinium* was positively correlated with native vegetation in the catchment and negatively correlated to TN and water column DOC. The two abundant

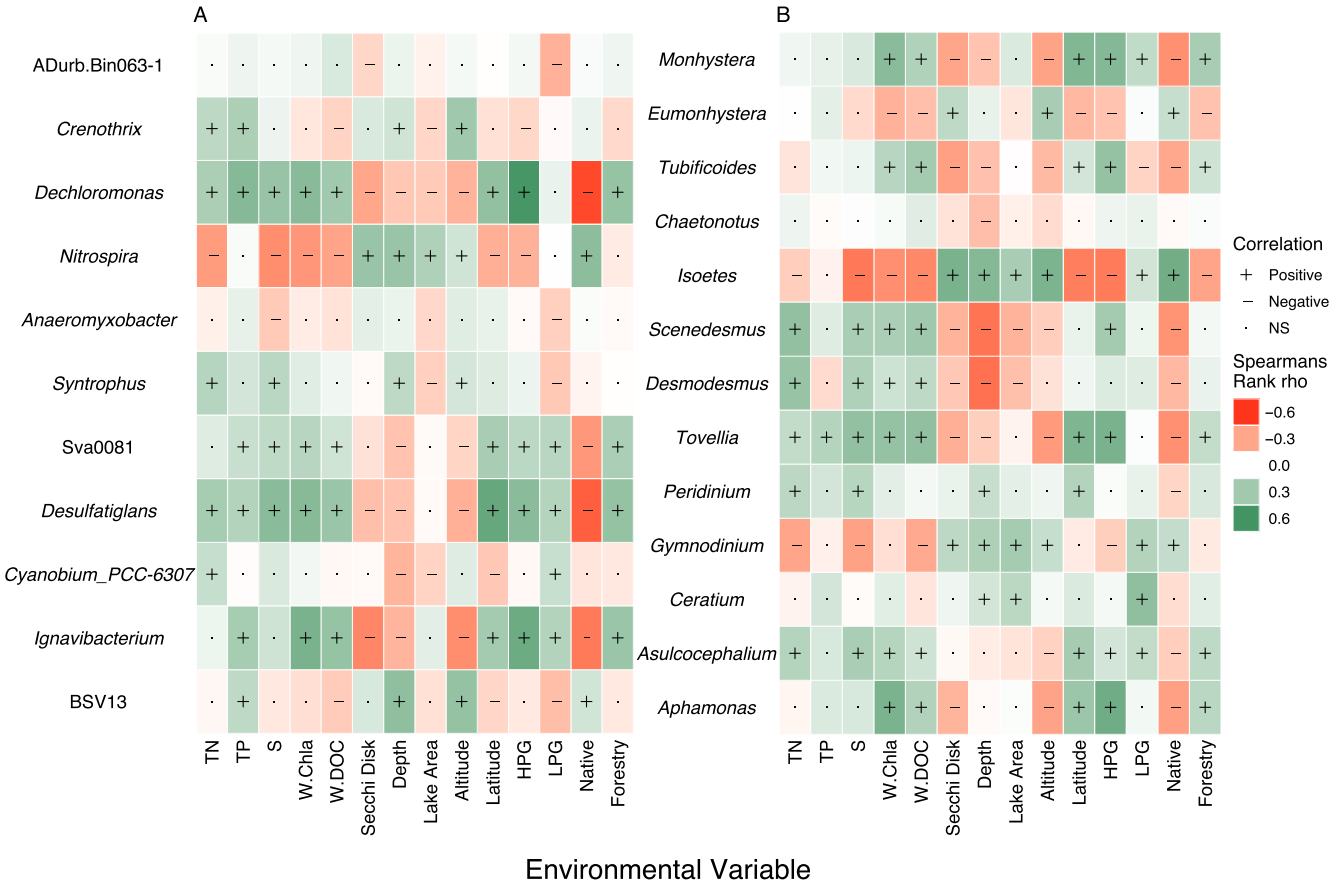

Fig. 7. Spearman's rank correlations for abundant (relative abundance greater than 1.5% in at least 10% of lakes) genera against environmental variables. Non-significant correlations (NS) are based on a *p*-value of ≥ 0.05. Alt, lake altitude; Area, lake area; Dep, maximum lake depth; HPG, high-productivity grassland in catchment; Lat, latitude; LPG, low-productivity grassland in catchment; Nat, native vegetation in catchment; S.D., Secchi disk depth; W.Chla, water column Chl *a*; W.DOC, water column dissolved organic carbon.

nematode genera showed contrasting patterns. *Monhystera* was positively associated with latitude and high-productivity grassland in the catchment as well as water column DOC and Chl *a* with *Eumonhystera* showing the opposite pattern (Figs. 6C,D, 7B; Supporting Information Table S4).

Discussion

This study investigated bacterial and eukaryotic assemblages in lake surface sediment from 296 lakes across 12° of latitude and a wide range of environmental gradients from high altitude glacial lakes in the South Island which are surrounded by a native catchment to nutrient rich lowland dune and swamp lakes in the North Island. To our knowledge, this is the largest investigation of bacterial and eukaryotic assemblages in lake surface sediments to date. The scale of the study and diversity of lakes included enabled us to investigate processes involved in assembling sediment assemblages and environmental factors driving taxa abundance across multiple interacting environmental gradients.

Overall diversity and occupancy

The finding that even across a large spatial scale Proteobacteria, especially Gammaproteobacteria, were the dominant bacteria within lake sediments both in terms of the number of ASVs and relative proportion of reads indicates this taxon plays an important role in lake sediment ecology. It has been observed previously to be major contributor for a geographically restricted subset of these data (Pearman et al. 2020) and is also consistent with studies on individual or a few lakes in other countries (Huang et al. 2017; Zhang et al. 2019, 2020). Gammaproteobacteria are a diverse phylum that play a vital role in degradation and metabolism in sediments (Huang et al. 2017).

Among eukaryotes, nationwide patterns were similar to those previously observed in the Southern lakes of New Zealand (Pearman et al. 2020) with Dinophyceae and Chlorophyceae high in ASVs, and the classes Clitellata and Ostracoda being ASV poor but accounting for a higher proportion of reads. Dinophyceae have been shown to be the substantial component in terms of reads in Lake Baikal (Yi et al. 2017). Likewise, Dinophyceae and Chlorophyceae comprised a sizeable proportion of the phytoplankton

community in Qinghai–Tibetan lakes (Liu et al. 2016). Dinophyceae and Chlorophyceae along with the prokaryotic Cyanobacteria are likely to play a substantial role in the primary production within these lakes. The presence of these phototrophs in the DNA of the sediment could be due to either them growing on the sediment substrate, being present in the form of resting stages or having settled out of the water column. Further experiments are required to investigate the origin of the phototroph DNA and these should also include exploring the impact of lake depth on the rate at which planktonic DNA settles in the sediment.

Positive abundance–occupancy relationships have been noted for a wide range of taxa from bacteria to vertebrates (Shade et al. 2018). The results presented here suggest that both the bacterial and eukaryotic assemblages in the surface sediments of lakes on a national scale follow the same pattern, a trend also observed for the bacterioplankton in a subset of these lakes (Pearman et al. 2022*a*). The majority of the taxa found in this study where present only in one lake and had very low relative abundances. The restricted distribution of rare organisms has been shown in a range of taxa, but the advent of molecular methods has highlighted the high proportion of rare microbes (Pedrós-Alió 2012). As microbes are easily dispersed a proportion of the rare microbes could have originated in surrounding environments (e.g., terrestrial habitats) and been transported into the lake where they can no longer propagate (Pedrós-Alió 2012). Abundance–occupancy theory suggests that taxa that have a wide niche breadth can tolerate a wider range of environmental conditions and thus have a higher occupancy rate and abundance (Izabel-Shen et al. 2021). The bacterial ASV found in the most lakes belonged to the genus Sva0081, a sulfate reducing bacteria which has been shown previously to play important roles in the sediments of lakes dominated by cyanobacteria (Fan et al. 2018). However, its presence even in lakes with high water quality and without cyanobacteria suggests that it could contribute substantially to sulfur cycling within lake sediments.

Among eukaryotes an ASV belonging to the class Cryptomycota was present in the most lakes. These fungi are parasites of a variety of eukaryotic organisms (Gleason et al. 2012) and have previously been found in lake sediments (Simon et al. 2016) and may be involved in transferring energy from primary producers to tertiary consumers (Gleason et al. 2012). The eukaryotes exhibited a weaker abundance–occupancy relationship with many ASVs having a restricted distribution and a relative high abundance. Several of these ASVs belonged to the phylum Dinoflagellata potentially providing an example of conditional rarity where taxa are rare until environmental conditions are suitable and a rapid increase in abundance is observed (Shade et al. 2014). Some of the ASVs that were abundant but restricted in distribution belonged to larger mobile organisms such those in the class Clitellata. Macroorganisms tend to have a lower density in sediments and thus a patchier distribution in the sediment and 0.25 g sediment samples are not suitable for the investigation of these organisms (Pawlowski et al. 2021). However, when collected in the sample due to their size they will likely account for high read counts which could led to the deviation in the abundance–occupancy relationships. Further studies using different sample volumes and analysis methods are required to explore drivers of assemblage composition for larger mobile eukaryotes in lakes.

DNA in the sediment can originate from both organisms that lived in the water column and have settled to the bottom or from those who inhabit the sediment. Due to limitations in the resolution of the taxonomic classifications, as well as a lack of knowledge of niche preferences for taxa, especially those attributed to environmental clades, this study did not attempt to differentiate between these groups of organisms. The incorporation of DNA from dead organisms may have slightly impacted the assessment of sediment assemblages although it is representative of the assemblage in and around the lake. Future studies could undertake metabarcoding on the RNA in the sediment as this is likely to be more reflective of the active assemblage.

The role of deterministic verses stochastic processes in driving assembly

In both bacterial and eukaryotic assemblages as well as the microbial subset of eukaryotes, deterministic processes were inferred to contribute more to assembly than stochastic processes. The dominance of deterministic over stochastic processes has been observed previously in lake sediments (Zeng et al. 2019) but contrasts with large scale planktonic bacterial survey in Canada (Kraemer et al. 2020). In the present study, variable selection was inferred to be the main process governing the assembly of the bacterial assemblage and an important contributor for the eukaryotes alongside homogeneous selection. Given the wide range of environmental gradients (e.g., altitude, lake depth, lake size, trophic state, catchment land use) included in this study, we anticipated that selective processes would contribute substantially to structuring the sediment assemblages. Variation partitioning indicated that land use in the catchment, which is strongly related to trophic state, was the most important environmental factor in explaining the assemblage composition, and a strong gradient between lakes in catchments with high productive grassland and those with native catchments was noted in the dbRDA for both bacteria and eukaryotes. This is in agreement with Kraemer et al. (2020) who have shown a pervasive signal of land use on bacterioplankton communities. Investigation of the assembly processes showed that for variable selection, 85% of the lake pairwise comparisons for both bacteria and eukaryotes were between trophic states indicating that nutrient concentrations and algal concentrations were exerting sufficient environmental pressure on taxa to drive selective adaptation. This agrees with Zeng et al. (2019) who

showed that beta diversity increased with greater differences in trophic state across 10 Chinese freshwater lakes in Nanjing, China. Such variable selection pressures likely underpin the success of using microbial indicator species for predicting trophic state (Pearman et al. 2022*b*).

Homogeneous selection was inferred to be the dominant assembly process in approximately a quarter of the pairwise comparisons for both the bacteria and the eukaryotes and ~ 37% for the microbial component of the eukaryotes. Homogeneous selection has previously been shown to be less important than variable selection for bacteria in lake sediments compared to those within the water column and is likely due to the more diverse habitat niches being present in the sediment (Jiao et al. 2020, 2021). The higher levels of homogeneous selection compared to variable selection for the microbial eukaryotes is in contrast to the bacteria and larger eukaryotes. Dinoflagellates were a major component of the microbial eukaryotes. These DNA sequences may have originated from the water column before settling into the sediment which could explain the higher values of homogeneous selection, which would be in agreement with the results of Jiao et al. (2020, 2021). Further work, which distinguishes living and dead organisms, is required to better understand these assembly patterns.

Stochastic processes were inferred to be less important compared to deterministic ones for both bacteria and eukaryotes. Dispersal limitation was slightly higher for eukaryotes than bacteria and agreed with the distribution patterns observed which were more restricted for eukaryotic ASVs. With no hydrological connection between the vast majority of lakes in this study (~ 5% had a connection to another lake) dispersal vectors are likely to be limited. Waterbirds have been shown to move zooplankton and dinoflagellates between lakes (Figuerola et al. 2005; Tesson et al. 2018) and could be a possible vector for long distance dispersal of both bacteria and eukaryotes between lakes. While modeling has shown aerial dispersal is viable for bacteria, long distance dispersal of organisms greater than 20 μm is unlikely (Wilkinson et al. 2012). It has been shown in the plankton that larger organisms such as zooplankton have higher dispersal limitations compared to phytoplankton and bacteria (Soininen et al. 2011). Similar size-related dispersal patterns may occur in the sediment explaining the higher levels of dispersal limitations of eukaryotes compared to bacteria. This is especially true as dispersal limitation of the microbial fraction of the eukaryotes is lower than that of eukaryotes overall suggesting the possible presence of size limitations on dispersal.

The contribution of homogenizing dispersal to the assembly of the surface sediment bacterial and eukaryotic assemblages was inferred to be low. Lakes can be considered aquatic islands surrounded by inhospitable terrestrial habitats; therefore, dispersal rates would not be expected to be sufficient to lead to homogenizing dispersal especially when lakes were not positioned close to one another. Zeng et al. (2019) showed that homogenizing dispersal did not contribute substantially to bacterial assemblages in surface sediments of 10 Chinese freshwater lakes across a trophic gradient. The authors also demonstrated that bacterial assemblages in the sediment were less likely to be predominantly assembled by homogenizing dispersal than planktonic assemblages indicating lower dispersal capabilities of sediment assemblages.

Environmental drivers of assemblage composition

Overall dbRDA analysis indicated that the measured environmental variables only explained a relatively small percentage of the variation in bacterial and eukaryotic assemblages. This result was surprising given the importance of deterministic processes in assembling both the bacterial and eukaryotic assemblages. There could be various reasons for this. First, dbRDA is a linear method and alternative approaches may identify nonlinear effects that are poorly identified using dbRDA. Second, environmental variables (e.g., anoxia, sediment temperature, pH, dissolved oxygen), not measured in this study can affect microbial assemblages as well as geological factors such as origin and age of the lake. Parameters such as these may explain for a portion of the remaining variation. Another explanation could be that biological interactions (e.g., grazing, commensalism, mutualism, and parasitism) have been shown to impact assemblage structure (Zhou and Ning 2017). While microbial interactions have been shown to exist, they are difficult to document, and little is known about how they impact assemblage composition (Nemergut et al. 2013).

Species within a genus can have substantially different responses to environmental variables, however, given the high number of ASV and to enable broad relationships to be identified we used correlations between the dominant genera and predictor variables in this study. Four genera of bacteria (*Dechloromonas*, Sva0081, *Desulfatiglans* and *Ignavibacterium*) were strongly associated with increased proportions of high-productivity grassland in the catchment as well as increased nutrients and DOC within the lake. *Dechloromonas* is an anaerobic nitrate reducer which has been shown to be an important constituent of wastewater treatment plants and contributor to the removal of nitrogen and phosphorus from these systems (Zhang et al. 2021). Desulfatiglans are sulfate reducing organisms that have been shown to be important in the sulfur cycle of sediments in aquatic systems (Jochum et al. 2018). They have also been shown to be positively correlated with TN in lakes sediment (Pan et al. 2020) in agreement with the results presented here.

Scendesmus and *Desmodesmus* are common green algae that often constitute a large proportion of algal biomass in freshwater systems and are often associated with nutrient rich conditions (An et al. 1999). This concurs with the distribution of these genera in the lakes of this study where they were present in higher abundance in higher trophic state lakes and had

positive correlations with TN in the sediment as well as water column dissolved oxygen and Chl *a* concentration.

Nematodes are important components of the meiofauna and perform important roles in benthic foodwebs of lakes (Ristau and Traunspurger 2011). The two dominant genera found in this study *Monhystera* and *Eumonhystera*, are both bacterial feeders (Ristau and Traunspurger 2011), but showed contrasting patterns in respect to environmental drivers. *Monhystera* was associated with more nutrient rich lakes with *Eumonhystera* being more abundant in low trophic states. This concurs with a study of Swedish lakes where species in the genus *Monhystera* were indicative of eutrophic lakes while the species *Eumonhystera longicaudatula* was more common in oligotrophic lakes (Ristau and Traunspurger 2011).

Conclusions

Large scale studies offer an unprecedented opportunity to enhance knowledge on ecological processes. Here we sampled surface sediments from 296 lakes which were representative of lakes found throughout New Zealand. These lakes crossed a variety of environmental gradients and provided an opportunity to investigate the processes structuring biological assemblages within lake surface sediments. We showed that bacteria and eukaryotes had a restricted distribution with the majority of ASVs present in one, or a few lakes. Our analysis, based on the Stegen framework, also indicated that deterministic processes, driven by environmental factors, contributed more than stochastic processes to structuring bacterial and eukaryotic assemblages. Variation partitioning indicated that land use in the catchment contributed was the most substantial individual component to explaining the composition of the sediment community. To our knowledge, this is the largest study to use the Stegen framework to investigate assembly processes on surface sediment assemblages in lakes. With increasing anthropogenic impact on lake systems, increased knowledge on the processes responsible for the structuring these assemblages are vital to assist efforts aimed at mitigating the impacts of humans and to allow lakes to recover from anthropogenic stressors.

Data availability statement

Raw sequence reads are available in the NCBI SRA archive under the accession numbers: PRJNA606991, PRJNA750120, and PRJNA813318.

References

Adrian, R., and others. 2009. Lakes as sentinels of climate change. Limnol. Oceanogr. **54**:2283–2297. doi:10.4319/lo.2009.54.6_part_2.2283

An, S. S., T. Friedl, and E. Hegewald. 1999. Phylogenetic relationships of *Scenedesmus* and *Scenedesmus*-like coccoid green algae as inferred from ITS-2 rDNA sequence comparisons. Plant Biol. **1**: 418–428. doi:10.1111/j.1438-8677.1999.tb00724.x

Bodenhofer, U., E. Bonatesta, C. Horejš-Kainrath, and S. Hochreiter. 2015. msa: An R package for multiple sequence alignment. Bioinformatics **31**: 3997–3999. doi:10.1093/bioinformatics/btv494

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods **13**: 581–583.

Capo, E., M.-E. Monchamp, M. J. L. Coolen, I. Domaizon, L. Armbrecht, and S. Bertilsson. 2022. Environmental paleomicrobiology: Using DNA preserved in aquatic sediments to its full potential. Environ. Microbiol. **24**: 2201–2209. doi:10.1111/1462-2920.15913

Carlson, R. E. 1977. A trophic state index for lakes. Limnol. Oceanogr. **22**: 361–369. doi:10.4319/lo.1977.22.2.0361

Compson, Z. G., B. McClenaghan, G. A. C. Singer, N. A. Fahner, and M. Hajibabaei. 2020. Front. Ecol. Evol.: 8:581835.

Dray, S., G. Blanchet, D. Borcard, G. Guenard, T. Jombart, P. Legendre, and H. H. Wagner. 2016. Adespatial: Multivariate multiscale spatial analysis. R package version 00-3 Httpscran R-Proj. Orgpackage Adespatial.

Dudgeon, D., and others. 2006. Freshwater biodiversity: Importance, threats, status and conservation challenges. Biol. Rev. **81**: 163–182. doi:10.1017/S1464793105006950

Duggan, I. C. 2007. An assessment of the water quality of ten Waikato lakes based on zooplankton community composition. CBER Contract Report 60. University of Waikato.

Fan, X., S. Ding, M. Gong, M. Chen, S. Gao, Z. Jin, and D. C. W. Tsang. 2018. Different influences of bacterial communities on Fe (III) reduction and phosphorus availability in sediments of the cyanobacteria- and macrophyte-dominated zones. Front. Microbiol. **9**:2636.

Figuerola, J., A. J. Green, and T. C. Michot. 2005. Invertebrate eggs can fly: Evidence of waterfowl-mediated gene flow in aquatic invertebrates. Am. Nat. **165**: 274–280. doi:10.1086/427092

Fine, P. V., and S. W. Kembel. 2011. Phylogenetic community structure and phylogenetic turnover across space and edaphic gradients in western Amazonian tree communities. Ecography **34**: 552–565.

Forsberg, C. 1989. Importance of sediments in understanding nutrient cyclings in lakes. Hydrobiologia **176**: 263–277. doi:10.1007/BF00026561

Gaston, K. J., T. M. Blackburn, J. J. D. Greenwood, R. D. Gregory, R. M. Quinn, and J. H. Lawton. 2000. Abundance–occupancy relationships. J. Appl. Ecol. **37**: 39–59. doi:10.1046/j.1365-2664.2000.00485.x

Gleason, F. H., L. T. Carney, O. Lilje, and S. L. Glockling. 2012. Ecological potentials of species of *Rozella* (Cryptomycota). Fungal Ecol. **5**: 651–656. doi:10.1016/j.funeco.2012.05.003

Guillou, L., and others. 2013. The protist ribosomal reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. **41**: D597–D604. doi:10.1093/nar/gks1160

Harrell Jr, F. E., and C. Dupont. 2021. Hmisc: Harrell miscellaneous. https://cran.r-project.org/web/packages/Hmisc/index.html

Herlemann, D. P., O. Geissinger, and A. Brune. 2007. The termite group I phylum is highly diverse and widespread in the environment. Appl. Environ. Microbiol. **73**: 6682–6685. doi:10.1128/AEM.00712-07

Hijmans, R. J., E. Williams, C. Vennes, and M. R. J. Hijmans. 2017. Package "geosphere". https://cran.r-project.org/web/packages/geosphere/index.html

Horikoshi, M., and others. 2022. ggfortify: Data visualization tools for statistical analysis results. https://cran.r-project.org/web/packages/ggfortify/index.html

Huang, W., X. Chen, X. Jiang, and B. Zheng. 2017. Characterization of sediment bacterial communities in plain lakes with different trophic statuses. Microbiol. Open **6**: e00503. doi:10.1002/mbo3.503

Irvine, F., L. C. Cwynar, J. C. Vermaire, and A. B. H. Rees. 2012. Midge-inferred temperature reconstructions and vegetation change over the last ~15,000 years from Trout Lake, northern Yukon Territory, eastern Beringia. J. Paleolimnol. **48**: 133–146. doi:10.1007/s10933-012-9612-7

Izabel-Shen, D., A.-L. Höger, and K. Jürgens. 2021. Abundance-occupancy relationships along taxonomic ranks reveal a consistency of niche differentiation in marine bacterioplankton with distinct lifestyles. Front. Microbiol. **12**:690712.

Ji, B., J. Liang, Y. Ma, L. Zhu, and Y. Liu. 2019. Bacterial community and eutrophic index analysis of the East Lake. Environ. Pollut. **252**: 682–688. doi:10.1016/j.envpol.2019.05.138

Jiao, C., D. Zhao, J. Zeng, L. Guo, and Z. Yu. 2020. Disentangling the seasonal co-occurrence patterns and ecological stochasticity of planktonic and benthic bacterial communities within multiple lakes. Sci. Total Environ. **740**: 140010. doi:10.1016/j.scitotenv.2020.140010

Jiao, C., D. Zhao, R. Huang, F. He, and Z. Yu. 2021. Habitats and seasons differentiate the assembly of bacterial communities along a trophic gradient of freshwater lakes. Freshw. Biol. **66**: 1515–1529. doi:10.1111/fwb.13735

Jochum, L. M., L. Schreiber, I. P. G. Marshall, B. B. Jørgensen, A. Schramm, and K. U. Kjeldsen. 2018. Single-cell genomics reveals a diverse metabolic potential of uncultivated *Desulfatiglans*-related Deltaproteobacteria widely distributed in marine sediment. Front. Microbiol. **9**: 2038. doi:10.3389/fmicb.2018.02038

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics **26**: 1463–1464. doi:10.1093/bioinformatics/btq166

Klindworth, A., E. Pruesse, T. Schweer, J. Peplies, C. Quast, M. Horn, and F. O. Gloeckner. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. **41**: e1. doi:10.1093/nar/gks808

Kraemer, S. A., N. Barbosa da Costa, B. J. Shapiro, M. Fradette, Y. Huot, and D. A. Walsh. 2020. A large-scale assessment of lakes reveals a pervasive signal of land use on bacterial communities. ISME J. **14**: 3011–3023. doi:10.1038/s41396-020-0733-0

Kuhn, M. 2008. Building predictive models in R using the caret package. J. Stat. Softw. **28**: 1–26.

Leathwick, J. R., D. West, D. Kelly, H. Robertson, D. Brown, W. L. Chadderton, and A.-G. Ausseil. 2010. Freshwater eEcosystems of New Zealand (FENZ) Geodatabase—version 1—user guide. Department of Conservation.

Li, H., J. Zeng, L. Ren, J. Wang, P. Xing, and Q. L. Wu. 2017. Contrasting patterns of diversity of abundant and rare bacterioplankton in freshwater lakes along an elevation gradient. Limnol. Oceanogr. **62**: 1570–1585. doi:10.1002/lno.10518

Li, Y., J. Meng, C. Zhang, S. Ji, Q. Kong, R. Wang, and J. Liu. 2020. Bottom-up and top-down effects on phytoplankton communities in two freshwater lakes. PLoS One **15**: e0231357. doi:10.1371/journal.pone.0231357

Liu, J., Z. Meng, X. Liu, and X.-H. Zhang. 2019. Microbial assembly, interaction, functioning, activity and diversification: A review derived from community compositional data. Mar. Life Sci. Technol. **1**: 112–128. doi:10.1007/s42995-019-00004-3

Liu, X., W. Hou, H. Dong, S. Wang, H. Jiang, G. Wu, J. Yang, and G. Li. 2016. Distribution and diversity of cyanobacteria and eukaryotic algae in Qinghai–Tibetan Lakes. Geomicrobiol. J. **33**: 860–869. doi:10.1080/01490451.2015.1120368

Logares, R., S. V. M. Tesson, B. Canbäck, M. Pontarp, K. Hedlund, and K. Rengefors. 2018. Contrasting prevalence of selection and drift in the community structuring of bacteria and microbial eukaryotes. Environ. Microbiol. **20**: 2231–2240. doi:10.1111/1462-2920.14265

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. **17**: 10–12. doi:10.14806/ej.17.1.200

McArdle, B. H., and M. J. Anderson. 2001. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. Ecology **82**: 290–297 doi: 10.1890/0012-9658(2001)082[0290:FMMTCD]2.0.CO;2.

McCrackin, M. L., H. P. Jones, P. C. Jones, and D. Moreno-Mateos. 2017. Recovery of lakes and coastal marine ecosystems from eutrophication: A global meta-analysis. Limnol. Oceanogr. **62**: 507–518. doi:10.1002/lno.10441

McMurdie, P. J., and S. Holmes. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One **8**: e61217. doi:10.1371/journal.pone.0061217

Nemergut, D. R., and others. 2013. Patterns and processes of microbial community assembly. Microbiol. Mol. Biol. Rev. **77**: 342–356. doi:10.1128/MMBR.00051-12

Ning, D., and others. 2020. A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. Nat. Commun. **11**: 4717. doi:10.1038/s41467-020-18560-z

Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, M. H. H. Stevens, M. J. Oksanen, and M. Suggests. 2007. The vegan package. Community ecology package, v. **10**. p. 631–637. https://cran.r-project.org/web/packages/vegan/index.html

Pan, X., L. Lin, H. Huang, and J. Chen. 2020. Differentiation of nitrogen and microbial community in the sediments from Lake Erhai, Yunnan–Kweichow Plateau, China. Geomicrobiol. J. **37**: 818–825. doi:10.1080/01490451.2020.1783034

Pawlowski, J., and others. 2021. Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. Sci. Total Environ. **818**: 151783. doi:10.1016/j.scitotenv.2021.151783

Pearman, J. K., and others. 2020. Local factors drive bacterial and microeukaryotic community composition in lake surface sediment collected across an altitudinal gradient. FEMS Microbiol. Ecol. **96**:fiaa070. doi:10.1093/femsec/fiaa070

Pearman, J. K., and others. 2022*a*. The role of environmental processes and geographic distance in regulating local and regionally abundant and rare bacterioplankton in lakes. Front. Microbiol. **12**:793441.

Pearman, J. K., and others. 2022*b*. A bacterial index to estimate lake trophic level: National scale validation. Sci. Total Environ. **812**: 152385. doi:10.1016/j.scitotenv.2021.152385

Pedrós-Alió, C. 2012. The rare bacterial biosphere. Ann. Rev. Mar. Sci. **4**: 449–466. doi:10.1146/annurev-marine-120710-100948

Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Gloeckner. 2007. SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. **35**: 7188–7196. doi:10.1093/nar/gkm864

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, p. 2014.

Ristau, K., and W. Traunspurger. 2011. Relation between nematode communities and trophic state in southern Swedish lakes. Hydrobiologia **663**: 121–133. doi:10.1007/s10750-010-0564-5

Rühland, K., A. Priesnitz, and J. P. Smol. 2003. Paleolimnological evidence from diatoms for recent environmental changes in 50 lakes across Canadian Arctic Treeline. Arct. Antarct. Alp. Res. **35**: 110–123. doi:10.1657/1523-0430(2003)035[0110:PEFDFR]2.0.CO;2

Ruuskanen, M. O., K. A. St. Pierre, V. L. St. Louis, S. Aris-Brosou, and A. J. Poulain. 2018. Physicochemical drivers of microbial community structure in sediments of Lake Hazen, Nunavut. Can. Front. Microbiol. **9:**1138. doi:10.3389/fmicb.2018.01138

Sadeghi, J., S. R. Chaganti, A. H. Shahraki, and D. D. Heath. 2021. Microbial community and abiotic effects on aquatic bacterial communities in north temperate lakes. Sci. Total Environ. **781**: 146771. doi:10.1016/j.scitotenv.2021.146771

Schallenberg, L. A., J. K. Pearman, C. W. Burns, and S. A. Wood. 2021. Metabarcoding reveals lacustrine picocyanobacteria respond to environmental change through adaptive community structuring. Front. Microbiol. **12**: 3258. doi:10.3389/fmicb.2021.757929

Schliep, K. P. 2011. phangorn: Phylogenetic analysis in R. Bioinformatics **27**: 592–593. doi:10.1093/bioinformatics/btq706

Shade, A., S. E. Jones, J. G. Caporaso, J. Handelsman, R. Knight, N. Fierer, and J. A. Gilbert. 2014. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. mBio **5**: 5–e01314. doi:10.1128/mBio.01371-14

Shade, A., and others. 2018. Macroecology to unite all life, large and small. Trends Ecol. Evol. **33**: 731–744. doi:10.1016/j.tree.2018.08.005

Simon, M., P. López-García, P. Deschamps, G. Restoux, P. Bertolino, D. Moreira, and L. Jardillier. 2016. Resilience of freshwater communities of small microbial eukaryotes undergoing severe drought events. Front. Microbiol. **7**: 812.

Sloan, W. T., M. Lunn, S. Woodcock, I. M. Head, S. Nee, and T. P. Curtis. 2006. Quantifying the roles of immigration and chance in shaping prokaryote community structure. Environ. Microbiol. **8**: 732–740. doi:10.1111/j.1462-2920.2005.00956.x

Soininen, J., J. J. Korhonen, J. Karhu, and A. Vetterli. 2011. Disentangling the spatial patterns in community composition of prokaryotic and eukaryotic lake plankton. Limnol. Oceanogr. **56**: 508–520. doi:10.4319/lo.2011.56.2.0508

Stegen, J. C., X. Lin, J. K. Fredrickson, X. Chen, D. W. Kennedy, C. J. Murray, M. L. Rockhold, and A. Konopka. 2013. Quantifying community assembly processes and identifying features that impose them. ISME J. **7**: 2069–2079.

Stegen, J. C., X. Lin, J. K. Fredrickson, and A. E. Konopka. 2015. Estimating and mapping ecological processes influencing microbial community assembly. Front. Microbiol. **6**: 370.

Tesson, S. V. M., A. Weißbach, A. Kremp, Å. Lindström, and K. Rengefors. 2018. The potential for dispersal of microalgal resting cysts by migratory birds. J. Phycol. **54**: 518–528. doi:10.1111/jpy.12756

Vellend, M. 2010. Conceptual synthesis in community ecology. Q. Rev. Biol. **85**: 183–206. doi:10.1086/652373

Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. **73**: 5261–5267.

Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. Springer.

Wilkinson, D. M., S. Koumoutsaris, E. A. D. Mitchell, and I. Bey. 2012. Modelling the effect of size on the aerial dispersal of microorganisms. J. Biogeogr. **39**: 89–97. doi:10.1111/j.1365-2699.2011.02569.x

Yi, Z., C. Berney, H. Hartikainen, S. Mahamdallie, M. Gardner, J. Boenigk, T. Cavalier-Smith, and D. Bass. 2017. High-throughput sequencing of microbial eukaryotes in Lake Baikal reveals ecologically differentiated communities and novel evolutionary radiations. FEMS Microbiol. Ecol. **93**. doi:10.1093/femsec/fix073

Zeng, J., C. Jiao, D. Zhao, H. Xu, R. Huang, X. Cao, Z. Yu, and Q. L. Wu. 2019. Patterns and assembly processes of planktonic and sedimentary bacterial community differ along a trophic gradient in freshwater lakes. Ecol. Indic. **106**: 105491. doi:10.1016/j.ecolind.2019.105491

Zhan, A., and others. 2013. High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. Methods Ecol. Evol. **4**: 558–565.

Zhang, L., T. Zhao, Q. Wang, L. Li, T. Shen, and G. Gao. 2019. Bacterial community composition in aquatic and sediment samples with spatiotemporal dynamics in large, shallow, eutrophic Lake Chaohu, China. J. Freshw. Ecol. **34**: 575–589. doi:10.1080/02705060.2019.1635536

Zhang, L., T. Shen, Y. Cheng, T. Zhao, L. Li, and P. Qi. 2020. Temporal and spatial variations in the bacterial community composition in Lake Bosten, a large, brackish lake in China. Sci. Rep. **10**: 304. doi:10.1038/s41598-019-57238-5

Zhang, S., C. Amanze, C. Sun, K. Zou, S. Fu, Y. Deng, X. Liu, and Y. Liang. 2021. Evolutionary, genomic, and biogeographic characterization of two novel xenobiotics-degrading strains affiliated with *Dechloromonas*. Heliyon **7**: e07181. doi:10.1016/j.heliyon.2021.e07181

Zhou, J., and D. Ning. 2017. Stochastic community assembly: Does it matter in microbial ecology? Microbiol. Mol. Biol. Rev. **81**: e00002–e00017. doi:10.1128/MMBR.00002-17

Acknowledgments

This research was funded by the New Zealand Ministry of Business, Innovation, and Employment research programme—Our lakes' health: past, present, future (Lakes380; C05X1707). J.K.P. thanks the Cawthron Internal Investment Fund for support. We thank all members of the Lakes380 team for field assistance (www.lakes380.com/about-the-project/the-team). The authors acknowledge the support of New Zealand regional authorities that provided data and permission to use it and assisted with access to the sampling sites: Northland Regional Council, Auckland Council, Waikato Regional Council, Bay of Plenty Regional Council, Hawkes Bay Regional Council, Taranaki District Council, Horizon Regional Council, Greater Wellington Regional Council, Marlborough District Council, Tasman District Council, West Coast Regional Council, Environment Canterbury, Otago Regional Council and Environment Southland. The authors thank iwi and landowners across the country for their assistance with sampling, accessing sites and guidance throughout this work. The Department of Conversation is acknowledged for assistance with permitting.

Conflict of Interest

None declared.

Submitted 31 March 2022
Revised 30 May 2022
Accepted 01 October 2022

Associate editor: Hans-Peter Grossart