Human access and deterministic processes play a major role in structuring planktonic and sedimentary bacterial and eukaryotic communities in lakes

John K Pearman [Corresp., 1], **Georgia Thomson-Laing** [1], **Lucy Thompson** [1], **Sean Waters** [1], **Marcus J Vandergoes** [2], **Jamie D Howarth** [3], **Ian C Duggan** [4], **Ian D Hogg** [4, 5], **Susanna A Wood** [1]

[1] Cawthron Institute, Nelson, New Zealand

[2] GNS Science, Lower Hutt, New Zealand

[3] Victoria University of Wellington, Wellington, New Zealand

[4] University of Waikato, Hamilton, New Zealand

[5] Canadian High Arctic Research Station, Nunavut, Canada

Corresponding Author: John K Pearman
Email address: John.Pearman@cawthron.org.nz

Metabarcoding Reveals Lacustrine Picocyanobacteria Respond to Environmental Change Through Adaptive Community Structuring Lakes provide habitat for a diverse array of species and offer a wide range of ecosystem services for humanity. However, they are highly vulnerable as they are not only impacted by adverse actions directly affecting them, but also those on the surrounding environment. Improving knowledge on the processes responsible for community assembly in different biotic components will aid in the protection and restoration of lakes. Studies to date suggested a combination of deterministic (where biotic/abiotic factors act on fitness differences amongst taxa) and stochastic (where dispersal plays a larger factor in community assembly) processes are responsible for structuring biotic communities, but there is no consensus on the relative roles these processes play, and data is lacking for lakes. In the present study, we sampled different biotic components in 34 lakes located on the South Island of New Zealand. To obtain a holistic view of assembly processes in lakes we used metabarcoding to investigate bacteria in the sediment and surface waters, and eukaryotes in the sediment and two different size fractions of the water column. Physicochemical parameters were collected in parallel. Results showed that deterministic processes dominated the assembly of lake communities although the relative importance of variable and homogeneous selection differed among the biotic components. Variable selection was more important in the sediment (SSbact and SSeuks) and for the bacterioplankton (Pbact) while the assembly of the eukaryotic plankton (SPeuks, LPeuks) was driven more by homogeneous selection. The ease of human access to the lakes had a significant effect on lake communities. In particular clade III of SAR11 and *Daphnia pulex*, were only present in lakes with public

access. This study provides insights into the distribution patterns of different biotic components and highlights the value in understanding the drivers of different biological communities within lakes.

1 **Human access and deterministic processes play a major role in structuring planktonic and**

2 **sedimentary bacterial and eukaryotic communities in lakes**

4 John K. Pearman[1], Georgia Thomson-Laing[1], Lucy Thompson[1], Sean Waters[1], Marcus J.

5 Vandergoes[2], Jamie D Howarth[3], Ian C. Duggan[4], Ian D. Hogg[4,5], Susanna A. Wood[1]

6 1. Cawthron Institute, Private Bag 2, Nelson 7042, New Zealand

7 2. GNS Science, PO Box 30-368, Lower Hutt 5040, New Zealand.

8 3. Victoria University of Wellington, PO Box 600, Wellington 6012, New Zealand

9 4. The University of Waikato, Private Bag 3105, Hamilton, New Zealand

10 5. Canadian High Arctic Research Station, PO Box 2150, Cambridge Bay, Nunavut, X0B 0C0, Canada

12 Corresponding author: John K Pearman; email: john.pearman@cawthron.org.nz

14 **Abstract**

15 Lakes provide habitat for a diverse array of species and offer a wide range of ecosystem services

16 for humanity. However, they are highly vulnerable as they are not only impacted by adverse

17 actions directly affecting them, but also those on the surrounding environment. Improving

18 knowledge on the processes responsible for community assembly in different biotic components

19 will aid in the protection and restoration of lakes. Studies to date suggested a combination of

20 deterministic (where biotic/abiotic factors act on fitness differences amongst taxa) and stochastic

21 (where dispersal plays a larger factor in community assembly) processes are responsible for

22 structuring biotic communities, but there is no consensus on the relative roles these processes

23 play, and data is lacking for lakes. In the present study, we sampled different biotic components

24 in 34 lakes located on the South Island of New Zealand. To obtain a holistic view of assembly

25 processes in lakes we used metabarcoding to investigate bacteria in the sediment and surface

26 waters, and eukaryotes in the sediment and two different size fractions of the water column.

27 Physicochemical parameters were collected in parallel. Results showed that deterministic

28 processes dominated the assembly of lake communities although the relative importance of

29 variable and homogeneous selection differed among the biotic components. Variable selection

30 was more important in the sediment (SSbact and SSeuks) and for the bacterioplankton (Pbact)

31 while the assembly of the eukaryotic plankton (SPeuks, LPeuks) was driven more by

32 homogeneous selection. The ease of human access to the lakes had a significant effect on lake

33   communities. In particular clade III of SAR11 and *Daphnia pulex*, were only present in lakes

34   with public access. This study provides insights into the distribution patterns of different biotic

35   components and highlights the value in understanding the drivers of different biological

36   communities within lakes.

38   **Introduction**

40   Lakes are often considered sentinels of environmental change as they accumulate nutrients and

41   environmental contaminants from the surrounding landscape and atmosphere (Williamson et al.,

42   2008; Adrian et al., 2009). The important and diverse ecosystem services provided by lakes (e.g.,

43   food and recreation) mean that humans have utilised lakes and their catchments for millennia.

44   Over this time, lakes have been subjected to multiple stressors including catchment land use

45   modification, increased nutrients, anthropogenic pollution and the introduction of non-

46   indigenous species (Dudgeon et al., 2006). These stressors have had profound effects on lake

47   biota and resulted in a global decline of lake ecosystem health (Arbuckle & Downing, 2001;

48   McCrackin et al., 2017). To ensure appropriate protection and limit the degradation in lake

49   health, it is vital to understand how biotic components are responding to stressors and how this

50   varies amongst lakes.

51

52   Lakes are ideal for studying biogeographic patterns as they are discrete bodies of water isolated

53   from each other by terrestrial habitats. Even over relatively small spatial scales, lakes can vary in

54   size, depth and geomorphology, and in the land use and geology of their surrounding catchments

55   (Wetzel, 2001). These inherent features of lakes provide an ideal opportunity for examining

56   differences in biotic communities and drivers of biodiversity among lakes. The water column and

57   sediments are two main habitat types within lakes which are characterised by distinctive

58   environmental properties, but which are in continuous contact and interconnected (Forsberg,

59   1989). Planktonic communities of lakes have received substantial research (Drakare & Liess,

60   2010; Newton et al., 2011; Liao et al., 2017; Yang & Zhang, 2020; Pearman et al., 2020;

61   Kraemer et al., 2020; Schallenberg et al., 2021; Duggan, Özkundakci & David, 2021). In

62   contrast, sediment communities have received comparatively less attention (Zeng et al., 2019;

63   Pearman et al., 2020; Jiao et al., 2020) despite sediments acting as a sink for organic matter,

64 nutrients and pollutants. Biological communities present within the sediments are also vital in

65 many biogeochemical processes (Forsberg, 1989). Studies investigating community patterns in

66 both the water column and sediments are notably fewer and focused mostly on differences in

67 bacterial communities between the two habitats (Zeng et al., 2019; Jiao et al., 2020, 2021; Xia et

68 al., 2022).

69

70 Various ecological theories have been developed to understand how communities are assembled.

71 Niche theory highlights the importance of deterministic processes and indicates that species can

72 co-exist by occupying different niches (Levine & HilleRisLambers, 2009). Hubbell's neutral

73 theory in contrast describes the important of stochastic processes such as birth/death and

74 colonization and extinction (Hubbell, 2001). Investigations of ecosystems have suggested a

75 combination of deterministic and stochastic assembly processes are responsible for structuring

76 biotic communities (Chase, 2010; Stegen et al., 2013; Zhou & Ning, 2017; Sadeghi et al., 2021),

77 however, there is no consensus on the relative roles these processes play (Vellend et al., 2014;

78 Stegen et al., 2015; Zhou & Ning, 2017). The basis of deterministic processes is that there are

79 differences in fitness amongst taxa that result in varied reproductive/replication rates (Vellend,

80 2010). Deterministic processes can be classified into either variable or homogeneous selection.

81 Variable selection occurs when environmental conditions vary spatially and / or temporally and

82 selective pressures result in high variation in the community structure (Zhou & Ning, 2017). In

83 contrast, homogeneous selection happens when there is little variation in the environmental

84 conditions leading to a more similar community structure than would be expected (Zhou & Ning,

85 2017). Previous studies have suggested that within lakes, abiotic factors such as pH, temperature

86 and nutrient conditions (Hoostal & Bouzat, 2008; Ruuskanen et al., 2018; Pearman et al., 2020)

87 and catchment characteristics, such as land use and altitude (Hayden & Beman, 2016; Li et al.,

88 2017; Pearman et al., 2020; Kraemer et al., 2020) can influence biotic communities. Further,

89 biotic interactions such as predation and competition can also have deterministic effects on

90 communities (Zhou & Ning, 2017). Stochastic processes can be relevant when taxa are

91 functionally equivalent and environmental factors are not strongly influencing community

92 assembly (Sloan et al., 2006; Drakare & Liess, 2010). Stochastic processes can be divided into

93 dispersal limitation and homogenizing dispersal. Dispersal limitation occurs when there are

94 barriers to the effective transfer of organisms between communities, coupled with ecological

95   drift leading to increased variations between communities (Stegen et al., 2015). In contrast,

96   homogenizing dispersal occurs when there is a high level of dispersal/gene flow between

97   communities resulting in communities that are more similar than would be expected by chance

98   alone (Zhou & Ning, 2017). Stochastic processes could be impacted by the access humans have

99   to lakes, with anthropogenic activities being implicated for spreading non-indigenous species

100   (Craig, 1992; Mills et al., 1994; Rowe, 2007; Kilroy et al., 2021).

101

102   Until recently, assembly processes have received limited attention in lakes and studies have

103   mainly been restricted to bacterial communities. In general, deterministic processes have been

104   shown to dominate the structuring of biotic communities (Liao et al., 2017; Logares et al., 2018;

105   Zeng et al., 2019; Jiao et al., 2020, 2021; Kraemer et al., 2020). In the water column,

106   homogeneous selection was found to be dominant for bacteria in research ranging from a large

107   scale study of Canadian lakes (Kraemer et al., 2020) to more restricted spatial scales in China

108   (Zeng et al., 2019; Jiao et al., 2020, 2021) and Antarctica (Logares et al., 2018). However, Jiao et

109   al., (2020) showed that assembly processes were seasonally variable. For eukaryotes, Logares et

110   al., (2018) could not identify a dominant process responsible for structuring communities in

111   Antarctic lakes, with the authors suggesting the community differences were a result of

112   ecological drift. They also highlighted that different biological taxonomic groups in the same

113   lakes could be structured by different processes. While investigations are limited, lake sediment

114   bacterial communities have been shown in a variety of Chinese lakes to be predominantly

115   assembled by variable selection. (Zeng et al., 2019; Jiao et al., 2020, 2021; Xia et al., 2022). In

116   general, there is a pattern of homogeneous selection in the water column and variable selection in

117   the sediment (Zeng et al., 2019; Jiao et al., 2020; Xia et al., 2022). This is most likely due to

118   variability in the sediment structure, as well as the accumulation of nutrients and pollutants from

119   the catchment making the sediment more heterogenous.

120

121   To enhance understanding of the distribution patterns of bacteria and eukaryotes in the water

122   column and sediment of lakes, we sampled different biotic components in 34 lakes located on the

123   South Island of New Zealand. We targeted bacteria in both the surface sediment and water, as

124   well as eukaryotes in the surface sediment and two different size fractions of eukaryotes in the

125   water column to obtain a more holistic view of assembly processes in lakes. The lakes sampled

126 varied along a gradient of human access enabling us to determine if this influenced biotic

127 communities. We hypothesised that: 1) each biotic component would have different assembly

128 processes; 2) communities in the sediment would have a higher contribution from variable

129 selection than those in the water column which would be predominantly assembled by

130 homogeneous selection; and 3) that lakes with human access would have different communities

131 to those with no access, because humans act as a mechanism for dispersal.

132

133 **Methods**

134

135 **Study lakes**

136 A total of 34 lakes were sampled on the central and western side of the South Island of New

137 Zealand between 21 January 2020 and 14 February 2020 (Fig. 1, Supplementary Table 1). Lakes

138 included a range of altitudes from 9 m (Lake Ianthe) to 1839 m (Duncan Stream 1) above sea

139 level.

140

141 Descriptions of the land cover in the catchments were derived from the most up to date satellite

142 imagery available in the Land Cover Database Version 5 (Landcare Research New Zealand Ltd

143 https://lris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-

144 zealand/). Eight broad groupings were used for the analysis and were: 1) native vegetation; 2)

145 urban; 3) non-indigenous vegetation; 4) water; 5) forestry; 6) high production grassland; 7) low

146 producing grassland; and 8) other (see Supplementary Table 1 for details of these groups).

147

148 Lakes were classified into four categories based on the ease of human access to the lake: 1) no

149 easy access, where the lakes were in a remote location with no public or formed access tracks; 2)

150 walking access, where there was a public walking/hiking track to the lake but no vehicle access;

151 3) private access, where access was possible by vehicle but restricted due to the lake being on

152 private land; and 4) public access, where the lake was publicly accessible by vehicle.

153

154 All samples were collected under the specifications of Special Permit 651 from the New Zealand

155 government agency Ministry for Primary Industries.

156

157 **Chlorophyll and nutrient analysis**

158 Surface water for chlorophyll *a* analysis was collected in triplicate sterile bottles (500 mL) from

159 a single site at the deepest point in the lake. A subsample (~ 200 mL from each bottle) was

160 filtered through a single GF/C filter (combining to ~ 600 mL) before being wrapped in

161 aluminium foil and stored at -20 °C. Analysis was undertaken at the Watercare Laboratories

162 (Auckland, New Zealand) following the APHA 10 200H method with a reporting limit of 0.0006

163 mg/L.

164

165 Water samples (1 L) were collected using an integrated tube sampled through the surface mixed

166 layer (as determined by RBR profiles, RBR Ltd, Ottawa, Canada). Total nutrient analysis was

167 undertaken on a flow injection analyzer using American Public Health Association (APHA)

168 4500 methods (https://www.standardmethods.org/). Reporting limits for total nitrogen (TN) and

169 total phosphorus (TP) analyses were 0.005 mg/L. Combustion analysis at 850 °C using APHA

170 5310 B methods (https://www.standardmethods.org/) was used for determination of total organic

171 carbon (TOC) with reporting limits of 0.5 mg/L.

172

173 Samples for analysis of sediment biogeochemistry were collected from a single site at the lake

174 depocenter. Three ponar grab samples were taken and sediment collected from the top two cm.

175 These were combined and stored in 500 g containers, which were chilled at 4°C and shipped to

176 the laboratory within 48 hrs. Samples were homogenized, centrifuged ($3,000 \times g$, 40 min, 4°C)

177 and the pore water decanted. Following the removal of carbonates by acid pre-treatment, TN and

178 TOC were analyzed using catalytic combustion at (900°C, O_2) and separation using a Thermal

179 Conductivity Detector (reporting limit for both g/100g sediment). Total P was measured in the

180 sediment by drying and sieving the sediment before acid digestion followed by Inductively

181 Coupled Plasma-Mass Spectrometry (ICP-MS) analysis based on the US Environmental

182 Protection Agency (EPA) method 200.8 with a reporting limit of 10 mg kg^{-1}.

183

184 **Biological sample collection and processing**

185 Surface water samples were collected in triplicate (500 mL) sterile bottles that had been rinsed

186 three times with lake water. Bacterioplankton (Pbact) were filtered onto a 0.22 µm S-Pak filter

187   (Millipore Sigma, USA) using a volume of 100 mL per replicate and stored at -20 °C until later

188   DNA extraction.

189

190   Small and large planktonic eukaryotes (SPeuks and LPeuks, respectively) were collected using a

191   20 and 40 μm mesh plankton net, respectively. Vertical trawls from ~ 1.5 m above the lakebed

192   were used except in Stoney Tarn which, due to its shallow depth, was sampled by pouring 10 L

193   of surface water through the nets. The concentrated biomass of each trawl was transferred to

194   separate 50 mL Falcon tubes and preserved with 95% ethanol.

195

196   A ponar grab sampler was used to collect surface sediment samples for both bacteria (SSbact)

197   and eukaryotes (SSeuks). Samples (1 g) were taken from the undisturbed surface layer (top 5

198   mm) ca using a sterile spatula and stored in LifeGuard™ Soil Preservation Solution (Qiagen,

199   USA) and stored at -80 °C for later DNA extraction.

200

201   Molecular analysis (i.e., DNA extraction, Polymerase Chain Reaction [PCR] set-up, template

202   addition and PCR analysis) were conducted in separate sterile laboratories dedicated to each step

203   to reduce potential cross contamination. Prior to DNA extraction, amplification set-up and

204   template addition the laboratories were UV sterilized for 15 mins. PCR set-up and template

205   addition were undertaken in laminar flow cabinets fitted with HEPA filters. For all steps aerosol

206   barrier tips (Eppendorf, Germany or Axygen, United States) were used.

207

208   DNA extractions were undertaken using the DNeasy PowerSoil Kit (Qiagen, Germany). For the

209   Pbact samples, filters were cut into five pieces using sterile techniques and placed in the

210   PowerBead tubes of the kit. For both the SPeuks and LPeuks samples, biomass was pelleted by

211   centrifugation ($3000 \times g$, 20 min, 4°C) and added to the PowerBead tubes. For the surface sediment

212   samples, a subsample (0.25 g) was placed in the PowerBead tubes. All further steps were as per

213   the DNeasy PowerSoil Kit instructions using a QIAcube. Negative extraction controls were

214   included after every 23 samples.

215

216   Bacteria, in the surface waters (Pbact) and surface sediment (SSbact) were targeted by primers

217   for the V3-V4 regions of the bacterial 16S ribosomal RNA gene (16S rRNA; Supplementary

218   Table S2; Herlemann, Geissinger & Brune, 2007; Klindworth et al., 2013). The SPeuks samples

219   as well as the SSeuks were amplified by primers for the 18S ribosomal RNA gene (18S rRNA;

220   Supplementary Table S2; Zhan et al., 2013). The LPeuks were assessed using mitochondrial

221   cytochrome oxidase I (COI) primers as this gene has good representation of zooplankton in

222   reference databases (Supplementary Table S2; Leray et al., 2013). Illumina™ overhang adaptors

223   were attached to the taxa specific primers to allow for dual-indexing (Kozich et al., 2013). PCR

224   reactions were undertaken in triplicate 20 µL volumes and contained 10 µL of MiFi 2 × PCR

225   mastermix, 2 µM of each primer (final conc.) and 2 µL of template DNA. All PCRs had an

226   initial denaturation step of 94 °C for 5 mins and a final extension time of 5 mins at 72 °C. The

227   cycling conditions for each primer are given in Supplementary Table 2. Negative controls for

228   both the PCR and extraction were included. SequalPrep normalization plates (ThermoFisher

229   Scientific, USA), were used to clean and normalize pooled triplicate PCR products. An Illumina

230   Miseq™ platform, at the Auckland Genomics Facility was used for sequencing. Sequence library

231   preparation was undertaken according to the Illumina 16S metagenomics library prep manual

232   (https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-

233   metagenomic-library-prep-guide-15044223-b.pdf) with the exception that after the indexing

234   PCR, 5 µL of each sample (including water samples acting as sequencing blank) was pooled and

235   a single clean-up was undertaken. Quality control on the library pool was assessed using a

236   bioanalyzer before dilution (to 4 nM), denaturation, and a further dilution to 7 ρM with a 15%

237   PhiX spike took place.

238

Bioinformatics

240   Bioinformatic pipelines for the three genes were identical unless otherwise stated. Cutadapt was

241   used to remove primer sequences from the raw reads with a maximum of one mismatch allowed

242   (Martin, 2011). Subsequent processing was undertaken using the DADA2 package (Callahan et

243   al., 2016) within R (R Core Team, 2020). For the 16S rRNA and 18S rRNA sequence reads were

244   truncated to 230 and 228 base pairs (bp) for forward and reverse reads respectively while for the

245   COI gene the truncation length was 312 bp for the forward and 314 for the reverse. Depending on

246   the quality of the sequencing run, sequences were filtered with a maximum number of "expected

247   errors" (maxEE) threshold of two or four for forward reads and four or six for reverse reads. Reads

248   that did not match this criterion were discarded and not used for further analysis. The inference of

249 Amplicon Sequence Variants (ASVs) was based on a parametric error matrix was constructed

250 based on the first 10^8 bp of the sequences. The paired-end reads were merged with a maximum

251 mismatch of 1 bp and a required minimum overlap of 10 bp before chimeric sequences were

252 removed using the removeBimeraDenovo script within the *DADA2* package.

253

254 For the COI dataset, pseudogenes were detected and removed using the methods described in

255 Leray and Knowlton (2015). Translated sequences were aligned against a subset of the MIDORI

256 database (Machida et al., 2017) using Multiple Alignment of Coding Sequences (MACSE; Ranwez

257 et al., 2011). Sequences were first assessed using the invertebrate translation code and then

258 subsequently the vertebrate code. Any sequences that contained a stop codon or greater than two

259 frame shifts for both translations were considered a pseudogene and removed from further

260 analysis.

261

262 The resulting chimera checked ASVs were used for taxonomic classification against the SILVA

263 138 (Pruesse et al., 2007) for the 16S rRNA, PR^2 (Guillou et al., 2013) for the 18S rRNA and a

264 combination of the BOLD (Ratnasingham & Hebert, 2007) and nucleotide NCBI (NCBI Resource

265 Coordinators, 2018) supplemented with some unpublished sequences from native zooplankton for

266 the COI gene datasets. Classification of sequences was undertaken using the rdp classifier (Wang

267 et al., 2007) with a bootstrap of 70 and 50 for the ribosomal and COI datasets respectively, to

268 enable classifications at higher taxonomic levels.

269

270 ASV tables can be found in Supplementary Table 3-7.

271

272 **Statistical Analysis**

273 The ASV table, taxonomy, and metadata were combined into a *phyloseq* object (McMurdie &

274 Holmes, 2013) for further analysis. Eukaryotic, chloroplast and mitochondrial ASVs were

275 removed from the 16S rRNA datasets (Pbact and SSbact), while bacteria were removed from both

276 the 18S rRNA (SPeuks and SSeuks) and COI (LPeuks) datasets. Contamination in the negative

277 controls was assessed and read numbers for each ASV found in the negative controls were

278 subtracted from the samples. The samples were subsampled to an even depth at a level of 24,134

279   reads for the SSbact, 23,188 for SSeuks, 11,710 for SPeuks, 25,297 for LPeuks and 51,362 for

280   Pbact.

281

282   Rarefaction curves were constructed using the number of reads sequenced across all the lakes

283   sampled for each biological component separately using the R package *ranacapa* (Kandlikar et al.,

284   2018). Alpha diversity metrics of the observed number of ASVs and Chao1 were calculated.

285

286   Assembly processes were assessed for the biological components of the lake using ecological

287   modelling frameworks developed and adapted by Stegen et al. (2013, 2015). Phylogenetic turnover

288   in the biological communities between pairs of lakes was assessed by calculating the mean-nearest-

289   taxon-distance (ßMNTD; Fine & Kembel, 2011) using the packages *msa* (Bodenhofer et al., 2015)

290   and *phangorn* (Schliep, 2011). The phylogenetic distance was quantified between each ASV in a

291   community and its nearest relative in a second community. A null distribution of ßMNTD was

292   calculated using the R packages *picante* (Kembel et al., 2010) and *iCAMP* (repetitions = 999, Ning

293   et al., 2020) using the assumption that ecological selection was not the primary cause of differences

294   in pairs of communities. The beta-nearest-taxon-index (ßMNTI) was calculated by comparing

295   between the null model and observed ßMNTD values and normalizing by standard deviation.

296   Deviations away from the null model indicated that deterministic processes were prevalent, which

297   were then allocated to one of two categories. 1) Homogenous selection, this will be dominant when

298   environmental conditions are similar leading to consistent selective pressures and resulting in low

299   levels of change in the community and ßMNTI values < -2. 2) Variable selection, occurs when

300   environmental differences are sufficient to assert selective pressures on the fitness differences

301   amongst taxa, resulting in higher than expected pairwise differences in communities and a ßMNTI

302   > 2. If pairwise comparisons did not deviate from the null distribution, then deterministic processes

303   were likely weak and stochastic processes were then evaluated. A stochastic null model was

304   calculated based on the Raup-Crick metric adapted to account for species relative abundances

305   (RC_{Bray}).The observed values were compared against the null model and standardised to between

306   -1 and 1 (Stegen et al., 2015). If large differences in the community are observed while

307   deterministic processes are low, then dispersal limitation and subsequent compositional drift of

308   communities is determined to be the predominant processes and is indicated by RC_{Bray} > 0.95.

309   Homogenizing dispersal, where high rates of dispersal led to similar communities is indicated

310   when deterministic processes are low and there are RC$_{Bray}$ values < -0.95. If there are no deviations

311   away from the null model for either the deterministic or stochastic processes, then the pairwise

312   comparison is interpreted as having no dominant assembly process. Assembly processes were

313   visualised using the R package *circlize* (Gu et al., 2014).

314

315   Distance similarity patterns between community similarity (1–Bray-Curtis) and environmental

316   distance were assessed with linear regression. Environmental variables consisted of nine lake

317   variables (TOC, TN, TP, chlorophyll *a* at the water surface, Secchi Disk depth, temperature just

318   above the sediment, dissolved oxygen just above the sediment, lake altitude, maximum depth of

319   the lake) and five catchment characteristics (proportion of forestry, native vegetation, non-

320   indigenous vegetation, low productivity grassland and other (predominantly gravel)). Values for

321   TOC, TP and TN were from the sediment for SSbact and SSeuks or from the water column for

322   Pbact, SPeuks and LPeuks. These variables were selected after checking for co-correlation and

323   non-zero variance using the packages *caret* (Kuhn, 2008) and *ggcorrplot* (Kassambara, 2019).

324   Selected variables were centred and normalised and missing values for Lake Mapourika were

325   imputed via bagging with the package *caret* before a distance matrix was constructed based on

326   Euclidean distances.

327

328   Multivariate differences in the community structure were assessed by a one-way permutational

329   analysis of variance (PERMANOVA; Anderson, 2017) with the factor Ease of human access (4

330   levels: No access, Walking, Private Access, Public Access). The multivariate analysis was based

331   on Bray-Curtis distance matrices constructed with the *vegan* library (Oksanen et al., 2007) in R

332   and undertaken with 999 permutations. Pairwise comparisons were undertaken using the package

333   *pairwiseAdonis* (Arbizu, 2020) with Benjamini-Hochberg adjusted p values. Differences in the

334   relative abundances of selected taxa depending on Access was undertaken using the Kruskal Wallis

335   test and pairwise comparisons were undertaken with the Dunn Test using Benjamini-Hochberg

336   adjusted p values. Environmental variables (sediment and water column TN, TP and TOC) were

337   tested for significant differences with human access pressure using a Kruskal-Wallis test and

338   Wilcoxon pairwise tests (Benjamini-Hochberg adjusted p values).

339

340   **Results**

341 ***Diversity and composition***

342 Rarefaction analysis showed that across all lakes, ASV richness reached a plateau for planktonic

343 bacteria (Pbact), surface sediment eukaryotes (SSeuks) and surface sediment bacteria (SSbact),

344 suggesting that the sampling effort was sufficient to capture the vast majority of gamma diversity

345 (Supplementary Fig. 1; Supplementary Table 8). The number of observed ASVs accounted for

346 greater than 97.5% as predicted by Chao1 for these communities (Supplementary Table 3). For

347 the planktonic eukaryotes (LPeuks and SPeuks) no plateau was reached indicating diversity

348 levels were underestimated although this was predominantly due to higher diversity in a couple

349 of lakes which did not reach saturation.

350

351 Sporichthyaceae (mean=12.0%, sd = 6.8%) and Comamonadaceae (mean = 10.9%, sd = 8.5%)

352 were the dominant components of the Pbact samples (Fig. 2). Copepods were the dominant

353 eukaryotic taxa in the water column with Centropagidae (mean = 39.2%, sd =42.7%) accounting

354 on average for the highest relative proportion of the LPeuks and Diaptomidae (mean= 43.3%,

355 sd=48.8%) for SPeuks. In the LPeuks other crustacean taxa, Daphniidae (mean = 23.4%, sd

356 =29.1%) and Bosminidae (mean = 10.1%, sd =20.8%) were also important components of the

357 community. In SPeuks, Cyclopidae (mean = 15.6%, sd = 26.4%) and the dinoflagellate

358 Ceratiaceae (mean = 10.3%, sd=21.6%) accounted for high relative proportions in a range of

359 lakes. The abundances of dominant taxa in the SPeuks and LPeuks were lake dependent with

360 generally a single dominant taxon (Supplementary Fig. 2). In the SSeuks community, *Ceratium*

361 *hirundinella* (mean 10.8%, sd= 24.5%) had the greatest average abundance although this was due

362 to high numbers occurring in a small number of lakes (Fig. 2 and Supplementary Fig. 2). The

363 bacterial family Bacteroidetes vadinHA17 had the highest average abundance across the lakes

364 (mean = 5.6%, sd = 4.0%) in the SSbact followed by Anaerolineaceae (mean = 5.3%, sd =4.3%)

365 and Pedophaeraceae (mean = 4.4%, sd =4.5%).

366

367 ***Assembly processes***

368

369 Overall, deterministic processes contributed more to the assembly of lake biotic communities

370 compared to stochastic processes (Table 1 and Fig. 3). Deterministic processes were highest in

371 the SSbact accounting for 81% of the comparisons. They were lowest in the SPeuks community

372 making up only about 27% of comparisons with most comparisons (54%) not being dominated
373 by a single process. For the sediment communities (both bacteria and eukaryotes) as well as the
374 Pbact, variable selection was more predominant compared with homogeneous selection while for
375 planktonic eukaryotic communities (LPeuks and SPeuks) the opposite was observed with
376 homogeneous selection contributing more. For the stochastic assembly processes in general,
377 dispersal limitation was more predominant except for SSbact where homogenizing dispersal
378 accounted for a slightly higher proportion of the lake pairwise comparisons. For all biotic
379 components there was a general tendency that lakes that were dominated by homogenizing
380 dispersal were closer together than those where dispersal limitation was predominant. However,
381 this difference was only significant for the Pbact (Kruskal Wallis: chi squared 7.94; $p = 0.005$)
382 and SSeuks (Kruskal Wallis: chi squared 13.92; $p < 0.001$).

383

384 Because deterministic processes were the major component of the assembly processes further
385 investigations of environmental distances versus community similarity were undertaken. The
386 bacterial communities showed a higher turnover (slope of the regression line) with increasing
387 dissimilarity in the environment especially in the surface sediments (Fig. 4). With the exception
388 of the SPeuks ($p = 0.68$), significant negative trends were present between community similarity
389 and environmental distance although r^2 values were low for all communities indicating a weak
390 trend.

391

392 ***Anthropogenic access pressures***

393

394 The environmental variables TN, TP and TOC in the water column and sediment were
395 investigated to assess if they varied with human access pressure. There was no significant
396 relationship between human access and sediment TN and TOC (Kruskall-Wallis chi sq = 6.5 and
397 5.1; p values = 0.091 and 0.164 respectively). However, there was a significant difference for
398 sediment TP ($p = 0.030$), although pairwise comparisons were not significant. In the water
399 column all variables were significantly different with human access pressure (TN: chi sq = 9.9; p
400 = 0.019. TP: chi sq = 18.3; $p < 0.001$. TOC chi sq = 13.8; p = 0.003.). In general, private lakes
401 had the highest levels of nutrients with pairwise comparisons showing significant differences
402 between private and no access lakes for all water column variables.

403

404   Multivariate analysis indicated that except for the SPeuks (p = 0.078), the community structure

405   of the biological components was significantly different depending on ease of human access

406   (Fig. 5). Pairwise comparisons indicated that differences were mainly related to the comparison

407   between lakes with no access and those with public access (Pbact p = 0.006; LPeuks p = 0.006;

408   SSeuks p = 0.006; SSbact p = 0.006). For the LPeuks (p = 0.006) there was a significant

409   difference between the communities with private access and those with no access and for the

410   Pbact (p = 0.021) and LPeuks (p = 0.023) there was a difference between lakes with private

411   access and public access.

412

413   Across all the biotic communities there were two taxa that showed strong patterns related to ease

414   of human access. In the Pbact, clade III of the class SAR11 had a significantly different relative

415   abundance (Kruskal-Wallis chi-squared = 15.66, p = 0.001; Table 2) in lakes depending on ease

416   of human access. It was not detected in the lakes with no access. It was present in 14 (out of 21)

417   of the lakes with access and had a significantly higher relative abundance in the lakes with public

418   access.

419

420   There were also significant differences (Kruskal-Wallis X^2 = 10.25, p = 0.016; Table 2) in the

421   invasive *Daphnia pulex* according to ease of human access with a higher average abundance in

422   the lakes with public access (mean = 29.2%; sd = 44.0%). Pairwise comparisons showed that

423   there was a significant difference between lakes with public access and those with no access (p =

424   0.01). *Daphnia pulex* was only detected in one of the 13 lakes which had no public access. The

425   lowest abundances were observed in the private access lakes where mean abundance was <

426   0.01%.

427

428   **Discussion**

429   Our analysis of 34 lakes on the South Island of New Zealand demonstrated that the assembly of

430   biotic communities in both lake water columns and bed sediments was primarily driven by

431   deterministic processes, although the contribution of the various processes differed between

432   biotic components. We also showed that the ease of human access to the lakes had a substantial

433   effect on the biotic communities in the lakes.

434

435 **Diversity and community composition**

436 The planktonic bacterial (Pbact) community had high proportions of Sporichthyaceae, within the

437 class Actinobacteria. This group is frequently found to be abundant within lakes across the world

438 (Glöckner et al., 2000; Warnecke, Amann & Pernthaler, 2004; Jiao et al., 2018; Kraemer et al.,

439 2020) and has previously been shown to be an important component in New Zealand lakes

440 (which included the lakes assessed here; Pearman et al., 2022). The small size of the

441 Actinobacteria, may protect them from protistan grazing (Allgaier & Grossart, 2006) which,

442 combined with their ability to degrade a variety of complex organic materials, makes them

443 competitive in lake environments (Ghai et al., 2014). The family Comamonadaceae, class

444 Gammaproteobacteria, was another taxa which contributed substantially to the Pbact and has

445 also previously been shown to be abundant in freshwater environments (Willems, 2014; Pearman

446 et al., 2022).

447

448 Zooplankton taxa, typical of New Zealand (Parkes & Duggan, 2012; Duggan, Özkundakci &

449 David, 2021), were found to dominate the large planktonic eukaryotes (LPeuks) in the study

450 lakes with the families Centropagidae, Daphniidae (comprised of the genus *Daphnia* and

451 *Ceriodaphnia*) and/or Bosminidae (genus *Bosmina*) especially prevalent. The dominance of

452 different zooplankton species in lakes is likely to have impacts on the food webs and carbon

453 cycling due to distinct size ranges of favoured prey with copepods (Centropagidae) preferring

454 larger sized prey compared to cladocerans (Daphniidae; Burns & Schallenberg, 1996; Sommer et

455 al., 2001). For example, lakes with Centropagidae in high relative abundances did not have high

456 relative abundances of Ceratiaceae (phytoplankton) in the SPeuks. In lakes not dominated by

457 copepods (e.g., those with high relative abundances of Daphniidae/Bosinidae), *Ceratium*

458 *hirundinella* had higher abudnaces. This might be due to lower top-down pressures controlling

459 its abundance thus allowing populations to increase, potentially forming blooms, as has been

460 suggested in Lake Hayes (New Zealand, Schallenberg & Schallenberg, 2017).

461

462 In the small planktonic eukaryotes (SPeuks), the copepod family Diaptomidae had the highest

463 relative abundance in most lakes. While Diaptomidae, which are non-indigenous, have been

464 reported in New Zealand lakes (Duggan, Green & Burger, 2006), they are not believed to be

465 widespread in lakes of the South Island. Given the correlation in occurrence between the

466 Diaptomidae in the SPeuks and Centropagidae in the LPeuks we investigated whether the

467 Diaptomidae had been misclassified. Further investigation indicated that there was no 18S rRNA

468 reference sequence belonging to the genus *Boeckella*, which was the dominant Centropagidae

469 taxa in these samples. Thus, this genus could not be correctly assigned using the 18S rRNA gene

470 if present in the SPeuks samples. The potential misclassification of ASVs belonging to the

471 family Diaptomidae highlights the issue of incomplete reference databases, potentially resulting

472 in the wrong conclusions being obtained. Further collaborations between taxonomists and

473 molecular biologists are required to improve reference databases so that they reflect the

474 communities being studied.

475

476 The family Ceratiaceae had the highest mean relative abundance in the SSeuk community as well

477 as the third highest in the SPeuks community with the ASVs taxonomically classified as

478 belonging to *Ceratium hirundinella*. This high abundance was mainly driven by the dominance

479 of this species in a restricted number of lakes. This mixotrophic dinoflagellate is known to form

480 problematic blooms in lakes which have been associated with fish kills in New Zealand lakes

481 (Bayer, Schallenberg & Martin, 2008; Schallenberg & Schallenberg, 2017). The high abundance

482 of this species, especially in Lakes Alexandrina, McGregor and Poerua, may indicate that blooms

483 are forming and potentially are impacting food-web interactions (Schallenberg & Schallenberg,

484 2017) and affecting ecosystem services.

485

486 In the SSbact community the abundances of the top ten taxa were more similar with

487 Anaerolineaceae and the environmental group Bacteroidetes vadinHA17 having the highest

488 average abundance. Both are considered to occupy anaerobic habitats and to be involved in the

489 degradation of complex organic matter (Wang et al., 2019).

490

491 **Community assembly**

492 The dominance of deterministic processes, especially variable selection, in both the SSbact and

493 SSeuks indicate that differences in the environmental variables present in the surface sediments

494 of the lakes are driving the assembly of these components. This agrees with various studies

495 undertaken on Chinese lakes (Zeng et al., 2019; Jiao et al., 2020, 2021). Sediment DNA includes

496   DNA from organisms living in the sediment, and those that have settled out from the water

497   column or been transported in from the surrounding catchment. The inclusion of these latter two

498   groups will impact the assessment of the processes dominating sediment assemblages. To assess

499   the component actively living in the sediment studies incorporating metabarcoding of the RNA

500   should be undertaken.

501

502   In the water column, deterministic processes were still dominant over stochastic processes

503   (although in SPeuks no dominant processes were most important). For the Pbact, variable

504   selection was the most important deterministic process. Other studies have suggested that

505   homogeneous selection is the dominant assembly process in the water column (Zeng et al., 2019;

506   Jiao et al., 2020; Kraemer et al., 2020) although Jiao et al. (2021) showed that dominance of

507   variable or homogeneous selection varies with season. Only a limited number of studies have

508   investigated the structuring of eukaryotic lake communities. In the present study, we found that

509   in general no dominant process was important in the structuring of SPeuks and LPeuks

510   communities. However, homogeneous selection dominated in some instances. Logares et al.,

511   (2018), in a study of eukaryotes in Antarctic lakes also showed that a substantial proportion of

512   the lakes had no dominant process (drift acting alone) structuring them. The prevalence of no

513   dominant processes may indicate that other factors such as initial inoculations, may be impacting

514   the structuring of the SPeuks and LPeuks communities.

515

516   Previous studies (Jiao et al., 2020, 2021; Xia et al., 2022) have shown that bacterial sediment

517   communities and water column communities were structured differently with homogeneous

518   selection being more prominent in the water column in comparison to the sediment. This pattern

519   was observed for both bacteria and eukaryotes although homogeneous selection was not the

520   dominant process for bacteria in the water column. This pattern suggests that the sediment

521   environment is more heterogeneous than the water column, providing more habitats for

522   organisms and allowing selection to drive more community divergence than would be expected

523   by chance. This is further exemplified with the current results by the turnover (slope of

524   community similarity against environmental distance) in communities being higher for sediment

525   communities compared to those in the water column.

526

527   In general, homogenizing dispersal contributed the least to the assembly of the biological

528   communities (except for the SSbact, where it was slightly higher than dispersal limitation). In the

529   current dataset this would be expected as the majority of the lakes studied were not connected

530   and thus dispersal would likely be via aerial means (Wilkinson et al., 2012) or by birds

531   (Figuerola, Green & Michot, 2005; Tesson et al., 2018), which may not be sufficiently great

532   enough to allow homogenizing dispersal to be substantially influential in the assembly of the

533   lake communities. It has previously been suggested that dispersal limitation is higher for larger

534   organisms such as zooplankton than phytoplankton and bacteria (Soininen et al., 2011).

535   However, in the current study this was not observed with dispersal limitation accounting for the

536   lowest proportion for LPeuks. However, it has been shown that dispersal limitation is scale

537   dependent (Havel & Shurin, 2004; Soininen et al., 2011), which could indicate that the scale of

538   the current study was not sufficient for dispersal limitation to be more important than local

539   environmental factors in the assembly of the biological communities.

540

541   **Anthropogenic access pressures**

542   The lakes in the current study have different levels of human access ranging from being nearly

543   inaccessible (except by helicopter or long off-track trekking) to having easy drivable public

544   access. In the present study, we showed that there was a significant difference in the lake

545   communities (except for the SPeuks) based on human access, with pairwise comparisons

546   indicating the differences were mainly between lakes with no access and those that were easily

547   accessible to the public. One possible reason for this could be differences in the environmental

548   variables based on human access. No clear trend was noted in the sediment geochemistry with

549   access pressure with only TP having a significant difference. Although pairwise comparisons

550   were not significant, in general lakes with private access had higher levels of TP with this likely

551   to be linked to the low productivity grassland in the catchment. In the water column significant

552   differences were noted for TN, TP and TOC with pairwise comparisons indicating that private

553   and public access lakes had higher values, this is most likely linked differences in the catchment.

554   Human alterations of the catchment have previously been shown to impact lake communities

555   (Kraemer et al., 2020). While not measured in this study increased human access would likely

556   result in higher concentrations of contaminants, such as polyaromatic hydrocarbons originating

557   from the combustion of fossil fuels by cars and boats. A small scale metagenomic study of lakes

558   in the South Island of New Zealand showed decreased levels of xenobiotic pathways in a lake

559   surrounded by a conservation zone compared to two lakes that had higher impacts from humans

560   (Biessy et al., 2022).

561   Noticeable differences were observed at the family level in the distribution of taxa dependent on

562   ease of human access. Most noticeably, the Pbact, clade III of SAR 11 was present in the

563   majority (70%) of the accessible lakes but completely absent from those lakes which were

564   classed as inaccessible. Clade III is a freshwater sister group of the marine group that dominate

565   marine waters and have been found to have a global distribution (Salcher, Pernthaler & Posch,

566   2011). While widely distributed, they are generally less abundant in the freshwater environments

567   than their marine counterparts (Newton et al., 2011) and thus, have received less scientific

568   interest. The lack of this group in any of the lakes that have no human access is interesting and

569   may indicate that for this group dispersal is naturally limited amongst lakes, with human

570   activities aiding dispersal. However, we acknowledge that other reasons such as the presence of

571   preferable environmental conditions could also explain some of the distribution patterns. While

572   little is known about the ecology of this clade seasonal patterns have been noted, with higher

573   abundance when water temperatures are greater than ~16 ºC (Salcher, Pernthaler & Posch,

574   2011). Most of the lakes with no clade III representation had surface waters with temperatures

575   less than ~16 ºC at the time of sampling suggesting that environmental conditions in these lakes

576   may limit the proliferation of this clade. Further work would be required to investigate the

577   environmental niche of this clade and to untangle the drivers that are limiting its distribution in

578   lake systems.

579   The proliferation of non-indigenous species within lake systems due to human activities has been

580   shown across the world and can have substantial impacts on the ecology of the lakes (Craig,

581   1992; Mills et al., 1994; Rowe, 2007; Kilroy et al., 2021). *Daphnia pulex* is not native to New

582   Zealand but has spread throughout the South Island of New Zealand since its introduction about

583   60 years ago (Ye et al., 2021) when it was thought to have been introduced with the stocking of

584   lakes with Chinook salmon (*Oncorhynchus tshawytschca*) (Duggan et al., 2012; Ye et al., 2021).

585   *Daphnia pulex* produces a higher number of ephippia compared to New Zealand native species.

586   Additionally, the ephippia of *D. pulex* float on the water surface for long periods of time until

587   they adhered to surfaces (Burns, 2013). The adherence to boats, fishing gear and other

588   recreational equipment could explain the higher prevalence of this species in the lakes with easy

589   public access. *Daphnia pulex* has the potential to considerably affect pelagic food webs of New

590   Zealand lakes, which are generally species poor (Burns, 2013).

591   **Conclusions**

592   In this study we used metabarcoding to investigate different biological components in 34 lakes

593   located on the South Island of New Zealand. The aim of our study was to enhance understanding

594   of how lake biotic communities are assembled and distributed. Our results showed that overall

595   deterministic processes dominated the assembly of lake communities. However, the contribution

596   of variable and homogeneous selection differed depending on the biotic component. Variable

597   selection was especially important in the sediment (SSbact and SSeuks) and for the

598   bacterioplankton (Pbact) while homogeneous selection was predominant for the eukaryotic

599   plankton (SPeuks, LPeuks). Our results also showed that human access impacts lake

600   communities. Two exemplars were clade III of SAR11 and *Daphnia pulex*, which were only

601   present in lakes with easy access. With increasing anthropogenic impacts on lakes, increased

602   knowledge on the distribution and assembly processes of the different ecological components in

603   lakes is vital in assisting efforts aimed at mitigating the effects humans have on these

604   ecosystems.

605

606   **Acknowledgements**

607   We thank all members of the Lakes380 team for field assistance (www.lakes380.com/about-the-

608   project/the-team). We acknowledge the support of New Zealand regional authorities that provided

609   data and permission to use it and assisted with access to the sampling sites: West Coast Regional

610   Council, Environment Canterbury and Otago Regional Council. The authors thank iwi and

611   landowners across the country for their assistance with sampling, accessing sites and guidance

612   throughout this work. The Department of Conversation is acknowledged for assistance with

613   permitting.

614   **References**

615   Adrian R, O'Reilly CM, Zagarese H, Baines SB, Hessen DO, Keller W, Livingstone DM,

616   Sommaruga R, Straile D, Van Donk E, Weyhenmeyer GA, Winder M. 2009. Lakes as

617   sentinels of climate change. *Limnology and Oceanography* 54:2283–2297. DOI:

618   10.4319/lo.2009.54.6_part_2.2283.

619   Allgaier M, Grossart H-P. 2006. Diversity and seasonal dynamics of Actinobacteria populations

620   in four lakes in northeastern Germany. *Applied and Environmental Microbiology*

621   72:3489–3497. DOI: 10.1128/AEM.72.5.3489-3497.2006.

622   Anderson MJ. 2017. Permutational Multivariate Analysis of Variance (PERMANOVA). In:

623   *Wiley StatsRef: Statistics Reference Online*. American Cancer Society, 1–15. DOI:

624   10.1002/9781118445112.stat07841.

625   Arbizu PM. 2020. pairwiseAdonis.

626   Arbuckle KE, Downing JA. 2001. The influence of watershed land use on lake N: P in a

627   predominantly agricultural landscape. *Limnology and Oceanography* 46:970–975. DOI:

628   10.4319/lo.2001.46.4.0970.

629   Bayer TK, Schallenberg M, Martin CE. 2008. Investigation of nutrient limitation status and

630   nutrient pathways in Lake Hayes, Otago, New Zealand: A case study for integrated lake

631   assessment. *New Zealand Journal of Marine and Freshwater Research* 42:285–295. DOI:

632   10.1080/00288330809509956.

633   Biessy L, Pearman JK, Waters S, Vandergoes MJ, Wood SA. 2022. Metagenomic insights to the

634   functional potential of sediment microbial communities in freshwater lakes.

635   *Metabarcoding and Metagenomics* 6:e79265. DOI: 10.3897/mbmg.6.79265.

636   Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. 2015. msa: an R package for

637   multiple sequence alignment. *Bioinformatics (Oxford, England)* 31:3997–3999. DOI:

638   10.1093/bioinformatics/btv494.

639   Burns CW. 2013. Predictors of invasion success by Daphnia species: influence of food,

640   temperature and species identity. *Biological Invasions* 15:859–869. DOI:

641   10.1007/s10530-012-0335-5.

642   Burns CW, Schallenberg M. 1996. Relative impacts of copepods, cladocerans and nutrients on

643   the microbial food web of a mesotrophic lake. *Journal of Plankton Research* 18:683–714.

644   DOI: 10.1093/plankt/18.5.683.

645   Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2:

646   high-resolution sample inference from Illumina amplicon data. *Nature Methods* 13:581.

647   Chase JM. 2010. Stochastic community assembly causes higher biodiversity in more productive

648   environments. *Science* 328:1388–1391.

649   Craig JF. 1992. Human-induced changes in the composition of fish communities in the African

650   Great Lakes. *Reviews in Fish Biology and Fisheries* 2:93–124. DOI:

651   10.1007/BF00042880.

652   Drakare S, Liess A. 2010. Local factors control the community composition of cyanobacteria in

653   lakes while heterotrophic bacteria follow a neutral model. *Freshwater Biology* 55:2447–

654   2457. DOI: 10.1111/j.1365-2427.2010.02473.x.

655   Dudgeon D, Arthington AH, Gessner MO, Kawabata Z-I, Knowler DJ, Lévêque C, Naiman RJ,

656   Prieur-Richard A-H, Soto D, Stiassny MLJ, Sullivan CA. 2006. Freshwater biodiversity:

657   importance, threats, status and conservation challenges. *Biological Reviews* 81:163–182.

658   DOI: https://doi.org/10.1017/S1464793105006950.

659   Duggan IC, Green JD, Burger DF. 2006. First New Zealand records of three non-indigenous

660   Zooplankton species: Skistodiaptomus pallidus, Sinodiaptomus valkanovi, and Daphnia

661   dentifera. *New Zealand Journal of Marine and Freshwater Research* 40:561–569. DOI:

662   10.1080/00288330.2006.9517445.

663   Duggan IC, Özkundakci D, David BO. 2021. Long-term zooplankton composition data reveal

664   impacts of invasions on community composition in the Waikato lakes, New Zealand.

665   *Aquatic Ecology* 55:1127–1142. DOI: 10.1007/s10452-020-09803-8.

666   Duggan IC, Robinson KV, Burns CW, Banks JC, Hogg ID. 2012. Identifying invertebrate

667   invasions using morphological and molecular analyses: North American Daphnia

668   'pulex'in New Zealand fresh waters. *Aquatic invasions* 7:585–590.

669   Figuerola J, Green AJ, Michot TC. 2005. Invertebrate eggs can fly: Evidence of

670   waterfowl-mediated gene flow in aquatic invertebrates. *The American Naturalist*

671   165:274–280. DOI: 10.1086/427092.

672   Fine PV, Kembel SW. 2011. Phylogenetic community structure and phylogenetic turnover across

673   space and edaphic gradients in western Amazonian tree communities. *Ecography*

674   34:552–565.

675   Forsberg C. 1989. Importance of sediments in understanding nutrient cyclings in lakes.

676   *Hydrobiologia* 176:263–277. DOI: 10.1007/BF00026561.

677   Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F. 2014. Key roles for freshwater

678   Actinobacteria revealed by deep metagenomic sequencing. *Molecular Ecology* 23:6073–

679   6090. DOI: 10.1111/mec.12985.

680   Glöckner FO, Zaichikov E, Belkova N, Denissova L, Pernthaler J, Pernthaler A, Amann R. 2000.

681   Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed

682   phylogenetic clusters including an abundant group of actinobacteria. *Applied and*

683   *Environmental Microbiology* 66:5053–5065. DOI: 10.1128/aem.66.11.5053-5065.2000.

684   Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. circlize implements and enhances circular

685   visualization in R. *Bioinformatics* 30:2811–2812. DOI: 10.1093/bioinformatics/btu393.

686   Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C,

687   Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra

688   WHCF, Lara E, Bescott NL, Logares R, Mahe F, Massana R, Montresor M, Morard R,

689   Not F, Pawlowski J, Probert I, Sauvadet A-L, Siano R, Stoeck T, Vaulot D, Zimmermann

690   P, Christen R. 2013. The Protist Ribosomal Reference database (PR2): a catalog of

691   unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic*

692   *Acids Research* 41:D597–D604. DOI: 10.1093/nar/gks1160.

693   Havel JE, Shurin JB. 2004. Mechanisms, effects, and scales of dispersal in freshwater

694   zooplankton. *Limnology and Oceanography* 49:1229–1238. DOI:

695   10.4319/lo.2004.49.4_part_2.1229.

696   Hayden CJ, Beman JM. 2016. Microbial diversity and community structure along a lake

697   elevation gradient in Yosemite National Park, California, USA. *Environmental*

698   *Microbiology* 18:1782–1791. DOI: 10.1111/1462-2920.12938.

699   Herlemann DP, Geissinger O, Brune A. 2007. The termite group I phylum is highly diverse and

700   widespread in the environment. *Applied Environmental Microbiology* 73:6682–5. DOI:

701   10.1128/AEM.00712-07.

702   Hoostal MJ, Bouzat JL. 2008. The modulating role of dissolved organic matter on spatial

703   patterns of microbial metabolism in Lake Erie sediments. *Microbial Ecology* 55:358–368.

704   DOI: 10.1007/s00248-007-9281-7.

705   Hubbell SP. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton

706   University Press.

707   Jiao C, Zhao D, Huang R, Cao X, Zeng J, Lin Y, Zhao W. 2018. Abundant and rare

708   bacterioplankton in freshwater lakes subjected to different levels of tourism disturbances.

709   *Water* 10:1075. DOI: 10.3390/w10081075.

710   Jiao C, Zhao D, Huang R, He F, Yu Z. 2021. Habitats and seasons differentiate the assembly of

711   bacterial communities along a trophic gradient of freshwater lakes. *Freshwater Biology*

712   66:1515–1529. DOI: 10.1111/fwb.13735.

713   Jiao C, Zhao D, Zeng J, Guo L, Yu Z. 2020. Disentangling the seasonal co-occurrence patterns

714   and ecological stochasticity of planktonic and benthic bacterial communities within

715   multiple lakes. *Science of The Total Environment* 740:140010. DOI:

716   10.1016/j.scitotenv.2020.140010.

717   Kandlikar GS, Gold ZJ, Cowen MC, Meyer RS, Freise AC, Kraft NJB, Moberg-Parker J,

718   Sprague J, Kushner DJ, Curd EE. 2018. ranacapa: An R package and Shiny web app to

719   explore environmental DNA data with exploratory statistics and interactive

720   visualizations. *F1000Research* 7. DOI: 10.12688/f1000research.16680.1.

721   Kassambara A. 2019. Package 'ggcorrplot.' *R package version 0.1* 3.

722   Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP,

723   Webb CO. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*

724   26:1463–1464. DOI: 10.1093/bioinformatics/btq166.

725   Kilroy C, Whitehead AL, Wood SA, Vandergoes MJ, Lambert P, Novis PM. 2021. Predicting

726   the potential distribution of the invasive freshwater diatom Lindavia intermedia in New

727   Zealand lakes. *Aquatic Invasions* 16.

728   Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Gloeckner FO. 2013.

729   Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-

730   generation sequencing-based diversity studies. *Nucleic Acids Research* 41. DOI:

731   10.1093/nar/gks808.

732   Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development of a dual-

733   index sequencing strategy and curation pipeline for analyzing amplicon sequence data on

734   the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*

735   79:5112–5120. DOI: 10.1128/AEM.01043-13.

736   Kraemer SA, Barbosa da Costa N, Shapiro BJ, Fradette M, Huot Y, Walsh DA. 2020. A large-

737   scale assessment of lakes reveals a pervasive signal of land use on bacterial communities.

738   *The ISME Journal* 14:3011–3023. DOI: 10.1038/s41396-020-0733-0.

739   Kuhn M. 2008. Building predictive models in R using the caret package. *Journal of Statistical*

740   *Software* 28:1–26.

741   Leray M, Knowlton N. 2015. DNA barcoding and metabarcoding of standardized samples reveal

742   patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences of*

743   *the United States of America* 112:2076–81. DOI: 10.1073/pnas.1424997112.

744 Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ. 2013.

745   A new versatile primer set targeting a short fragment of the mitochondrial COI region for

746   metabarcoding metazoan diversity: application for characterizing coral reef fish gut

747   contents. *Frontiers in Zoology* 10:34. DOI: 10.1186/1742-9994-10-34.

748 Levine JM, HilleRisLambers J. 2009. The importance of niches for the maintenance of species

749   diversity. *Nature* 461:254–257. DOI: 10.1038/nature08251.

750 Li H, Zeng J, Ren L, Wang J, Xing P, Wu QL. 2017. Contrasting patterns of diversity of

751   abundant and rare bacterioplankton in freshwater lakes along an elevation gradient.

752   *Limnology and Oceanography* 62:1570–1585. DOI: 10.1002/lno.10518.

753 Liao J, Cao X, Wang J, Zhao L, Sun J, Jiang D, Huang Y. 2017. Similar community assembly

754   mechanisms underlie similar biogeography of rare and abundant bacteria in lakes on

755   Yungui Plateau, China. *Limnology and Oceanography* 62:723–735. DOI:

756   10.1002/lno.10455.

757 Logares R, Tesson SVM, Canbäck B, Pontarp M, Hedlund K, Rengefors K. 2018. Contrasting

758   prevalence of selection and drift in the community structuring of bacteria and microbial

759   eukaryotes. *Environmental Microbiology* 20:2231–2240. DOI: 10.1111/1462-

760   2920.14265.

761 Machida RJ, Leray M, Ho S-L, Knowlton N. 2017. Metazoan mitochondrial gene sequence

762   reference datasets for taxonomic assignment of environmental samples. *Scientific Data*

763   4:1–7. DOI: 10.1038/sdata.2017.27.

764 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.

765   *EMBnet.journal* 17:10–12. DOI: 10.14806/ej.17.1.200.

766   McCrackin ML, Jones HP, Jones PC, Moreno-Mateos D. 2017. Recovery of lakes and coastal

767   marine ecosystems from eutrophication: A global meta-analysis. *Limnology and*

768   *Oceanography* 62:507–518. DOI: 10.1002/lno.10441.

769   McMurdie PJ, Holmes S. 2013. phyloseq: An R package for reproducible interactive analysis

770   and graphics of microbiome census data. *PLoS ONE* 8:e61217. DOI:

771   10.1371/journal.pone.0061217.

772   Mills EL, Leach JH, Carlton JT, Secor CL. 1994. Exotic species and the integrity of the Great

773   Lakes. *BioScience* 44:666–676. DOI: 10.2307/1312510.

774   NCBI Resource Coordinators. 2018. Database resources of the National Center for

775   Biotechnology Information. *Nucleic Acids Research* 46:D8–D13. DOI:

776   10.1093/nar/gkx1095.

777   Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. 2011. A guide to the natural history

778   of freshwater lake bacteria. *Microbiology and Molecular Biology Reviews* 75:14–49.

779   DOI: 10.1128/MMBR.00028-10.

780   Ning D, Yuan M, Wu L, Zhang Y, Guo X, Zhou X, Yang Y, Arkin AP, Firestone MK, Zhou J.

781   2020. A quantitative framework reveals ecological drivers of grassland microbial

782   community assembly in response to warming. *Nature Communications* 11:4717. DOI:

783   10.1038/s41467-020-18560-z.

784   Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M. 2007. The

785   vegan package. *Community ecology package* 10:631–637.

786   Parkes SM, Duggan IC. 2012. Are zooplankton invasions in constructed waters facilitated by

787   simple communities? *Diversity and Distributions* 18:1199–1210. DOI: 10.1111/j.1472-

788   4642.2012.00913.x.

789   Pearman JK, Biessy L, Thomson-Laing G, Waters S, Vandergoes MJ, Howarth JD, Rees A, Moy

790   C, Pochon X, Wood SA. 2020. Local factors drive bacterial and microeukaryotic

791   community composition in lake surface sediment collected across an altitudinal gradient.

792   *FEMS Microbiology Ecology* 96. DOI: 10.1093/femsec/fiaa070.

793   Pearman JK, Thomson-Laing G, Thomson-Laing J, Thompson L, Waters S, Reyes L, Howarth

794   JD, Vandergoes MJ, Wood SA. 2022. The role of environmental processes and

795   geographic distance in regulating local and regionally abundant and rare bacterioplankton

796   in lakes. *Frontiers in Microbiology* 12.

797   Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Gloeckner FO. 2007. SILVA: a

798   comprehensive online resource for quality checked and aligned ribosomal RNA sequence

799   data compatible with ARB. *Nucleic Acids Research* 35:7188–7196. DOI:

800   10.1093/nar/gkm864.

801   R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria:

802   R Foundation for Statistical Computing; 2014.

803   Ranwez V, Harispe S, Delsuc F, Douzert EJP. 2011. MACSE: Multiple Alignment of Coding

804   SEquences Accounting for Frameshifts and Stop Codons. *PLoS ONE*. DOI:

805   https://doi.org/10.1371/journal.pone.0022594.

806   Ratnasingham S, Hebert PDN. 2007. BOLD: The Barcode of Life Data System

807   (www.barcodinglife.org). *Molecular Ecology Notes* 7:355–364. DOI: 10.1111/j.1471-

808   8286.2006.01678.x.

809   Rowe DK. 2007. Exotic fish introductions and the decline of water clarity in small North Island,

810   New Zealand lakes: a multi-species problem. *Hydrobiologia* 583:345–358. DOI:

811   10.1007/s10750-007-0646-1.

812   Ruuskanen MO, St. Pierre KA, St. Louis VL, Aris-Brosou S, Poulain AJ. 2018. Physicochemical

813   drivers of microbial community structure in sediments of Lake Hazen, Nunavut, Canada.

814   *Frontiers in Microbiology* 9. DOI: 10.3389/fmicb.2018.01138.

815   Sadeghi J, Chaganti SR, Shahraki AH, Heath DD. 2021. Microbial community and abiotic

816   effects on aquatic bacterial communities in north temperate lakes. *Science of The Total*

817   *Environment* 781:146771. DOI: 10.1016/j.scitotenv.2021.146771.

818   Salcher MM, Pernthaler J, Posch T. 2011. Seasonal bloom dynamics and ecophysiology of the

819   freshwater sister clade of SAR11 bacteria 'that rule the waves' (LD12). *The ISME*

820   *Journal* 5:1242–1252. DOI: 10.1038/ismej.2011.8.

821   Schallenberg LA, Pearman JK, Burns CW, Wood SA. 2021. Metabarcoding reveals lacustrine

822   picocyanobacteria respond to environmental change through adaptive community

823   structuring. *Frontiers in Microbiology* 12:3258. DOI: 10.3389/fmicb.2021.757929.

824   Schallenberg M, Schallenberg LA. 2017. *Lake Hayes Restoration and Monitoring Plan*.

825   Dunedin: Hydrosphere Research Ltd.

826   Schliep KP. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics (Oxford, England)*

827   27:592–593. DOI: 10.1093/bioinformatics/btq706.

828   Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. 2006. Quantifying the roles of

829   immigration and chance in shaping prokaryote community structure. *Environmental*

830   *Microbiology* 8:732–740. DOI: 10.1111/j.1462-2920.2005.00956.x.

831   Soininen J, Korhonen JJ, Karhu J, Vetterli A. 2011. Disentangling the spatial patterns in

832   community composition of prokaryotic and eukaryotic lake plankton. *Limnology and*

833   *Oceanography* 56:508–520. DOI: https://doi.org/10.4319/lo.2011.56.2.0508.

834   Sommer U, Sommer F, Santer B, Jamieson C, Boersma M, Becker C, Hansen T. 2001.

835   Complementary impact of copepods and cladocerans on phytoplankton. *Ecology Letters*

836   4:545–550. DOI: 10.1046/j.1461-0248.2001.00263.x.

837   Stegen JC, Lin X, Fredrickson JK, Chen X, Kennedy DW, Murray CJ, Rockhold ML, Konopka

838   A. 2013. Quantifying community assembly processes and identifying features that

839   impose them. *The ISME journal* 7:2069.

840   Stegen JC, Lin X, Fredrickson JK, Konopka AE. 2015. Estimating and mapping ecological

841   processes influencing microbial community assembly. *Frontiers in Microbiology* 6:370.

842   Tesson SVM, Weißbach A, Kremp A, Lindström Å, Rengefors K. 2018. The potential for

843   dispersal of microalgal resting cysts by migratory birds. *Journal of Phycology* 54:518–

844   528. DOI: https://doi.org/10.1111/jpy.12756.

845   Vellend M. 2010. Conceptual synthesis in community ecology. *The Quarterly Review of Biology*

846   85:183–206. DOI: 10.1086/652373.

847   Vellend M, Srivastava DS, Anderson KM, Brown CD, Jankowski JE, Kleynhans EJ, Kraft NJB,

848   Letaw AD, Macdonald AAM, Maclean JE, Myers-Smith IH, Norris AR, Xue X. 2014.

849   Assessing the relative importance of neutral stochasticity in ecological communities.

850   *Oikos* 123:1420–1430. DOI: 10.1111/oik.01493.

851   Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment

852   of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental*

853   *Microbiology* 73:5261–5267.

854   Wang X-T, Xu X-J, Chen C, Xing D-F, Zhang R-C, Zhou X, Yuan Y, Wang A-J, Ren N-Q, Lee

855   D-J. 2019. The microbial zonation of SRB and soNRB enhanced the performance of SR-

856   DSR process under the micro-aerobic condition. *Environment International* 132:105096.

857   DOI: 10.1016/j.envint.2019.105096.

858   Warnecke F, Amann R, Pernthaler J. 2004. Actinobacterial 16S rRNA genes from freshwater

859   habitats cluster in four distinct lineages. *Environmental Microbiology* 6:242–253. DOI:

860   https://doi.org/10.1111/j.1462-2920.2004.00561.x.

861   Wetzel RG. 2001. *Limnology: Lake and River Ecosystems*. Gulf Professional Publishing.

862   Wilkinson DM, Koumoutsaris S, Mitchell EAD, Bey I. 2012. Modelling the effect of size on the

863   aerial dispersal of microorganisms. *Journal of Biogeography* 39:89–97. DOI:

864   10.1111/j.1365-2699.2011.02569.x.

865   Willems A. 2014. The family Comamonadaceae. In: *The prokaryotes: Alphaproteobacteria and*

866   *betaproteobacteria*. Springer, 777–851.

867   Williamson CE, Dodds W, Kratz TK, Palmer MA. 2008. Lakes and streams as sentinels of

868   environmental change in terrestrial and atmospheric processes. *Frontiers in Ecology and*

869   *the Environment* 6:247–254. DOI: https://doi.org/10.1890/070140.

870   Xia M, Xiong F, Li X, Li D, Wang Z, Zhai D, Liu H, Chen Y, Yu J, Wang Y. 2022. Different

871   assembly patterns of planktonic and sedimentary bacterial community in a few connected

872   eutrophic lakes. *Water* 14:723. DOI: 10.3390/w14050723.

873   Yang J, Zhang X. 2020. eDNA metabarcoding in zooplankton improves the ecological status

874   assessment of aquatic ecosystems. *Environment International* 134:105230. DOI:

875   10.1016/j.envint.2019.105230.

876   Ye Z, Williams E, Zhao C, Burns CW, Lynch M. 2021. The rapid, mass invasion of New

877   Zealand by North American Daphnia "pulex." *Limnology and Oceanography* 66:2672–

878   2683. DOI: 10.1002/lno.11780.

879 Zeng J, Jiao C, Zhao D, Xu H, Huang R, Cao X, Yu Z, Wu QL. 2019. Patterns and assembly

880   processes of planktonic and sedimentary bacterial community differ along a trophic

881   gradient in freshwater lakes. *Ecological Indicators* 106:105491. DOI:

882   10.1016/j.ecolind.2019.105491.

883 Zhan A, Hulák M, Sylvester F, Huang X, Adebayo AA, Abbott CL, Adamowicz SJ, Heath DD,

884   Cristescu ME, MacIsaac HJ. 2013. High sensitivity of 454 pyrosequencing for detection

885   of rare species in aquatic communities. *Methods in Ecology and Evolution* 4:558–565.

886 Zhou J, Ning D. 2017. Stochastic Community Assembly: Does It Matter in Microbial Ecology?

887   *Microbiology and Molecular Biology Reviews* 81:e00002-17. DOI:

888   10.1128/MMBR.00002-17.

889

890 **Legends:**

891 *Figure 1. Location of the lakes sampled. Details of sampling positions and lake characteristics*

892 *are given in Supplementary Table 1. Points are coloured based on the classification of ease of*

893 *human access to the lake.*

894

895 *Figure 2. Mean relative abundance of combined Amplicon Sequence Variants (ASVs) at the*

896 *family level. Error bars represent the standard deviation. Only the top 10 families are shown for*

897 *each community type. Note scale on y-axis varies among community plots.*

898

899 *Figure 3. Assembly processes for the different biological components. The width of the bar is*

900 *proportional to the percent of pairwise comparisons attributed to the assembly process. DL =*

901 *Dispersal limitation; HD = Homogenizing dispersal; VS = Variable selection; HS =*

902 *Homogeneous selection. NDP = No Dominant Process, Pbact = bacterioplankton; LPeuks =*

903 *large planktonic eukaryotes; SPeuks = small planktonic eukaryotes; SSeuks = surface sediment*

904 *eukaryotes; SSbact = surface sediment bacteria.*

905

906 *Figure 4. Community similarity (1 – Bray-Curtis) against environmental distance for the*

907 *various community components. The shaded grey area represents the 95% confidence intervals.*

908

909 *Figure 5: Non metric multidimensional scaling (NMDS) plots for the five biological components.*

910 *Points are coloured by the ease of human access to the lake.*

911

912 *Supplementary Figure 1. Rarefaction curves showing the number of Amplicon Sequence Variants*

913 *(ASVs) observed across the total sequencing effort for each biological component of the lakes.*

914 *Note the different values on x and y axes.*

915

916 *Supplementary Figure 2. Composition of the different biological components of each lake. Only*

917 *the top 10 classified families are shown for each community type.*

918

919 *Table 1. The proportion of lake pairwise comparisons attributed to each assembly process for*

920 *the different biological components assessed in the study. VS = variable selection HS =*

921 *homogeneous selection, DL = dispersal limitation, HD = homogenizing dispersal.*

922

923 *Table 2. The relative abundance and standard deviation of clade III of SAR11 and Daphnia*

924 *pulex depending on the ease of human access to the lakes.*

925

926 *Supplementary Table 1: Metadata of the 34 lakes sampled in this study.*

927

928 *Supplementary Table 2: Primer sequences and PCR conditions used in this study.*

929

930 *Supplementary Table 3: ASV table for the surface sediment bacteria (SSbact) component.*

931

932 *Supplementary Table 4: ASV table for the surface sediment eukaryote (SSeuks) component.*

933

934 *Supplementary Table 5: ASV table for the small planktonic eukaryote (SPeuks) component.*

935

936 *Supplementary Table 6: ASV table for the large planktonic eukaryote (LPeuks) component.*

937

938   *Supplementary Table 7: ASV table for the planktonic bacteria (Pbact) component.*

939
940   *Supplementary Table 8: Total number of Amplicon Sequence variants (ASVs) across the lakes*

941   *and the number of ASVs per lake.*

942

Figure 1

Location of the lakes sampled.

Details of sampling positions and lake characteristics are given in Supplementary Table 1.
Points are coloured based on the classification of ease of human access to the lake.

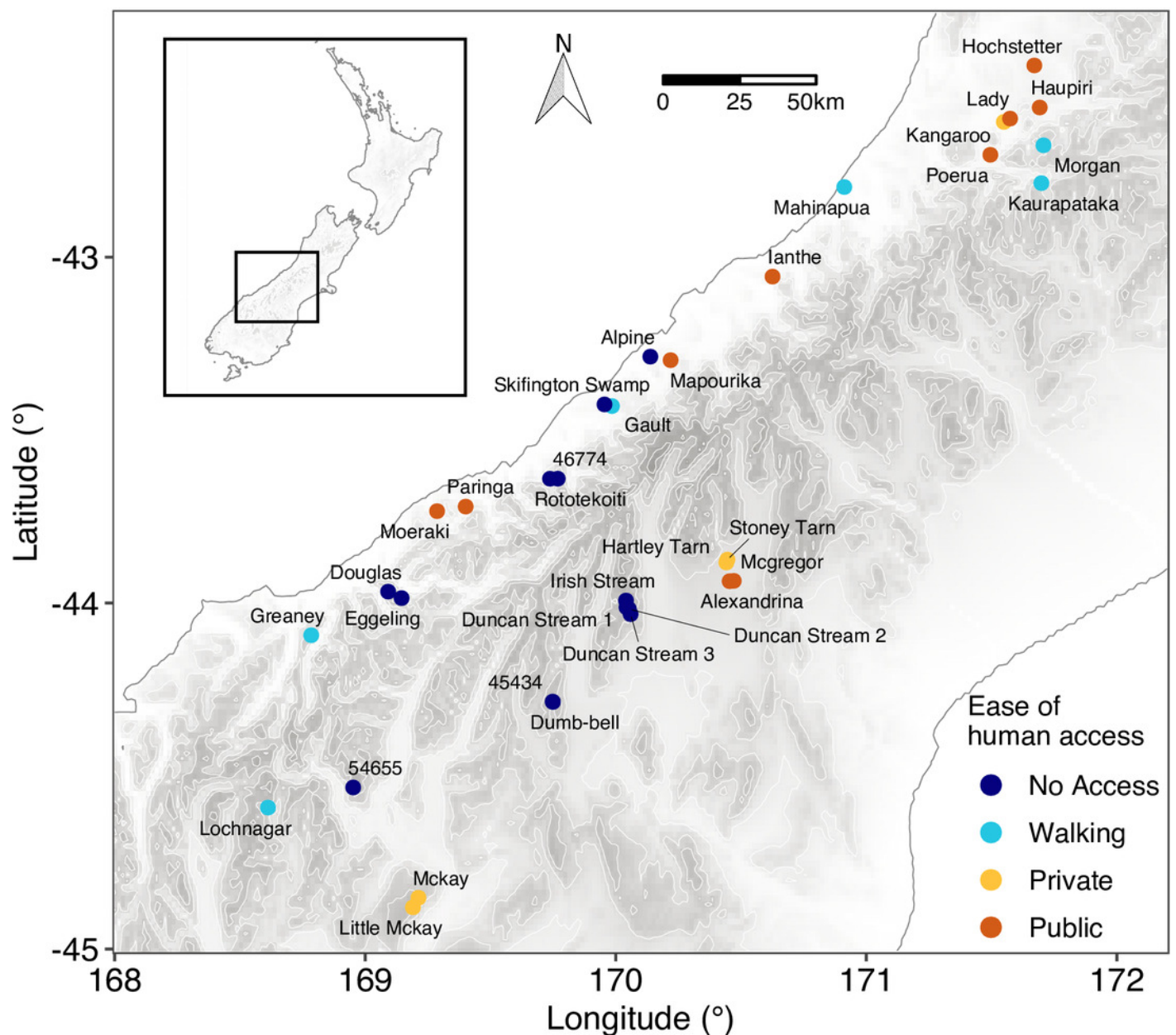



Figure 2

Mean relative abundance of combined Amplicon Sequence Variants (ASVs) at the family level.

Only the top 10 families are shown for each community type. Note scale on y-axis varies among community plots.

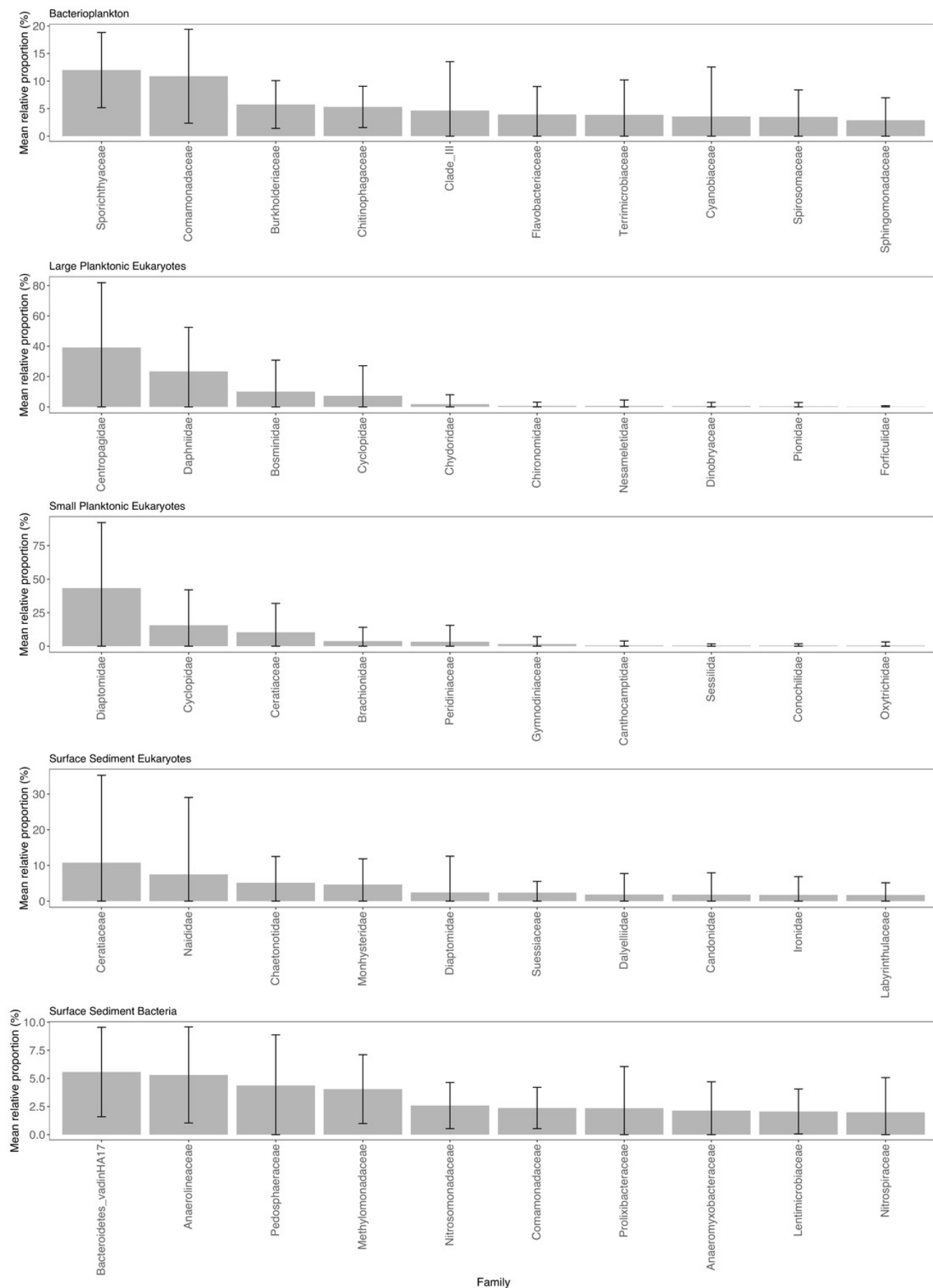



Figure 3

Assembly processes for the different biological components.

The width of the bar is proportional to the percent of pairwise comparisons attributed to the assembly process. DL = Dispersal limitation; HD = Homogenizing dispersal; VS = Variable selection; HS = Homogeneous selection. NDP = No Dominant Process, Pbact = bacterioplankton; LPeuks = large planktonic eukaryotes; SPeuks = small planktonic eukaryotes; SSeuks = surface sediment eukaryotes; SSbact = surface sediment bacteria.

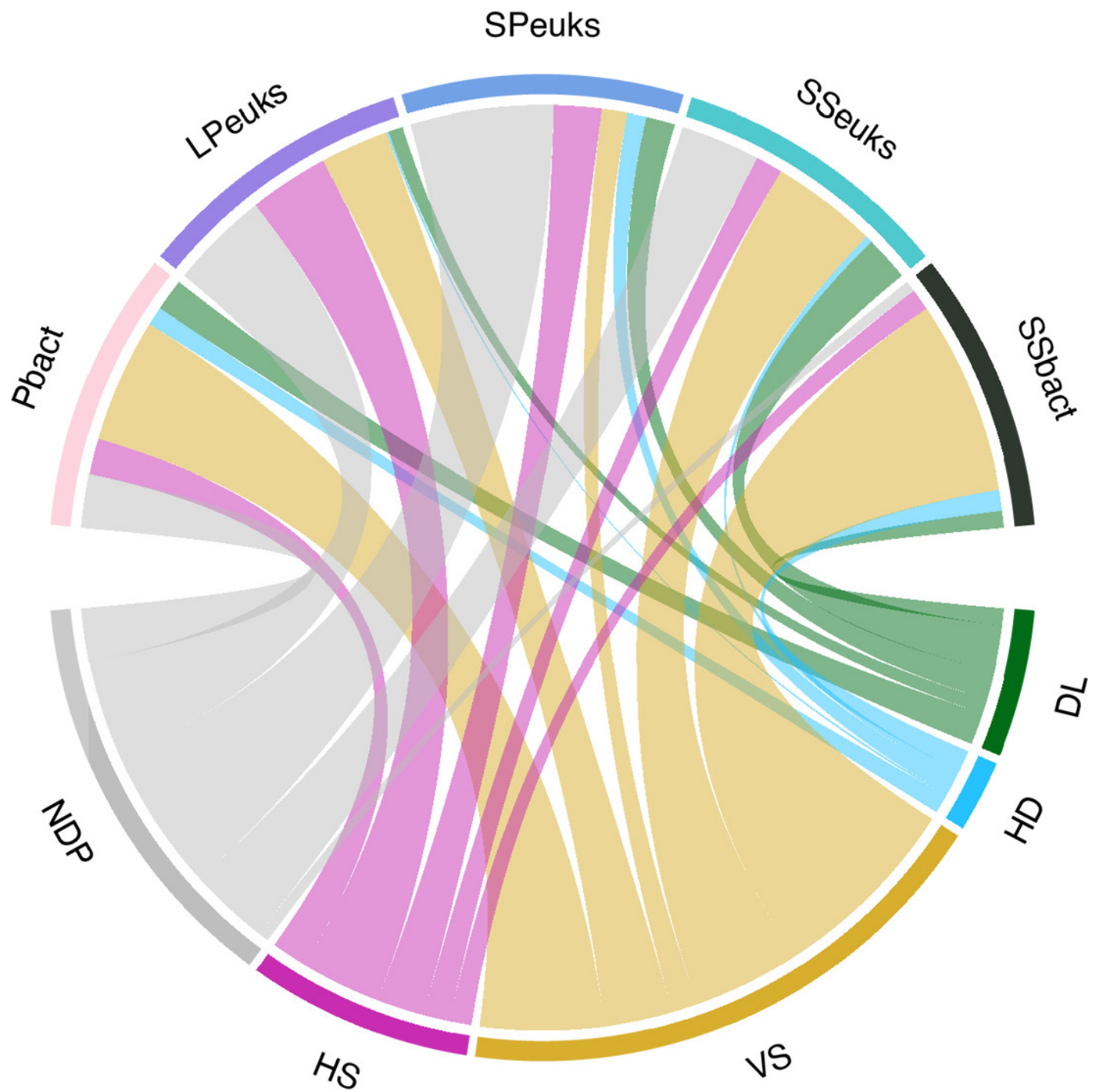



Figure 4

Community similarity (1 – Bray-Curtis) against environmental distance for the various community components.

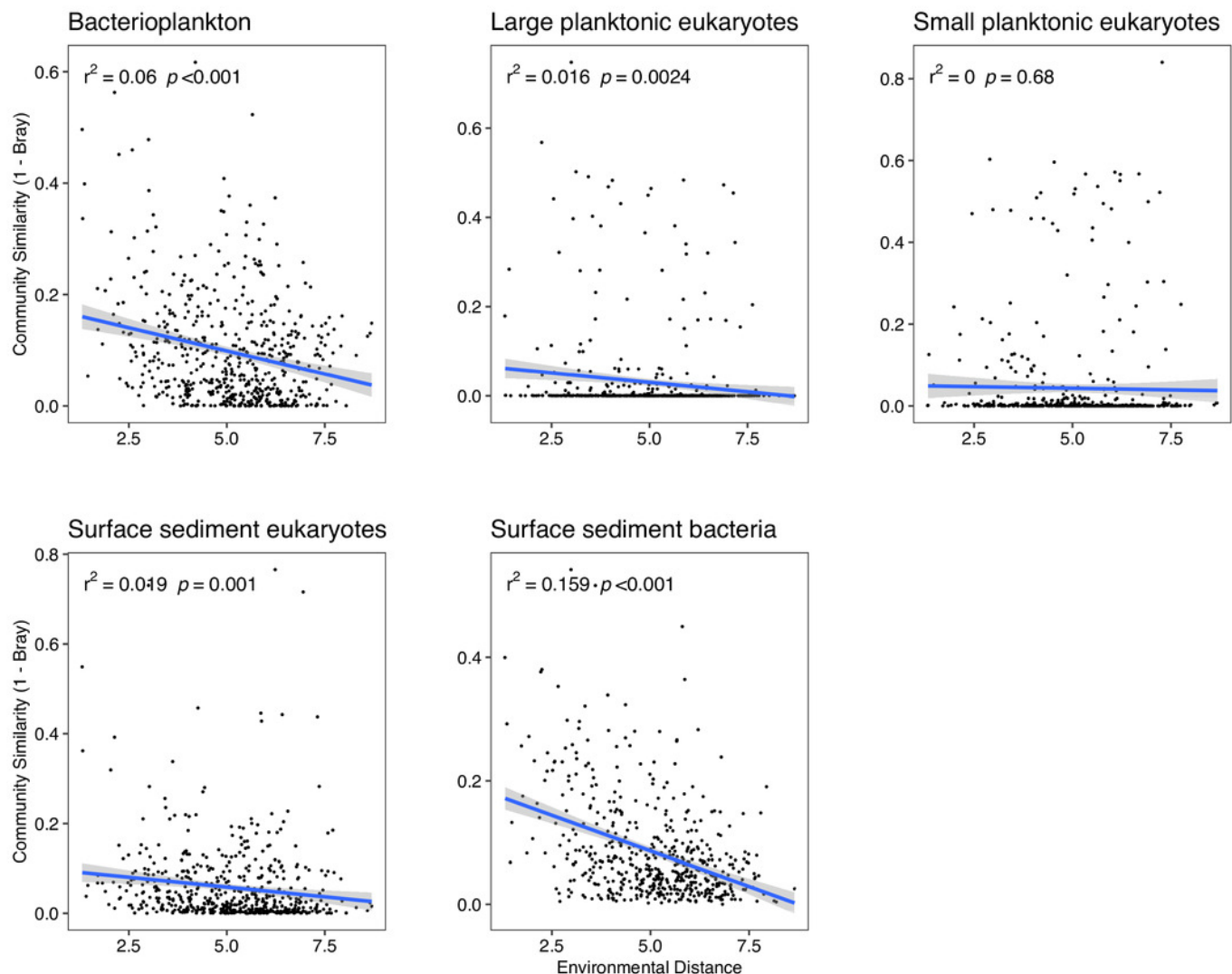
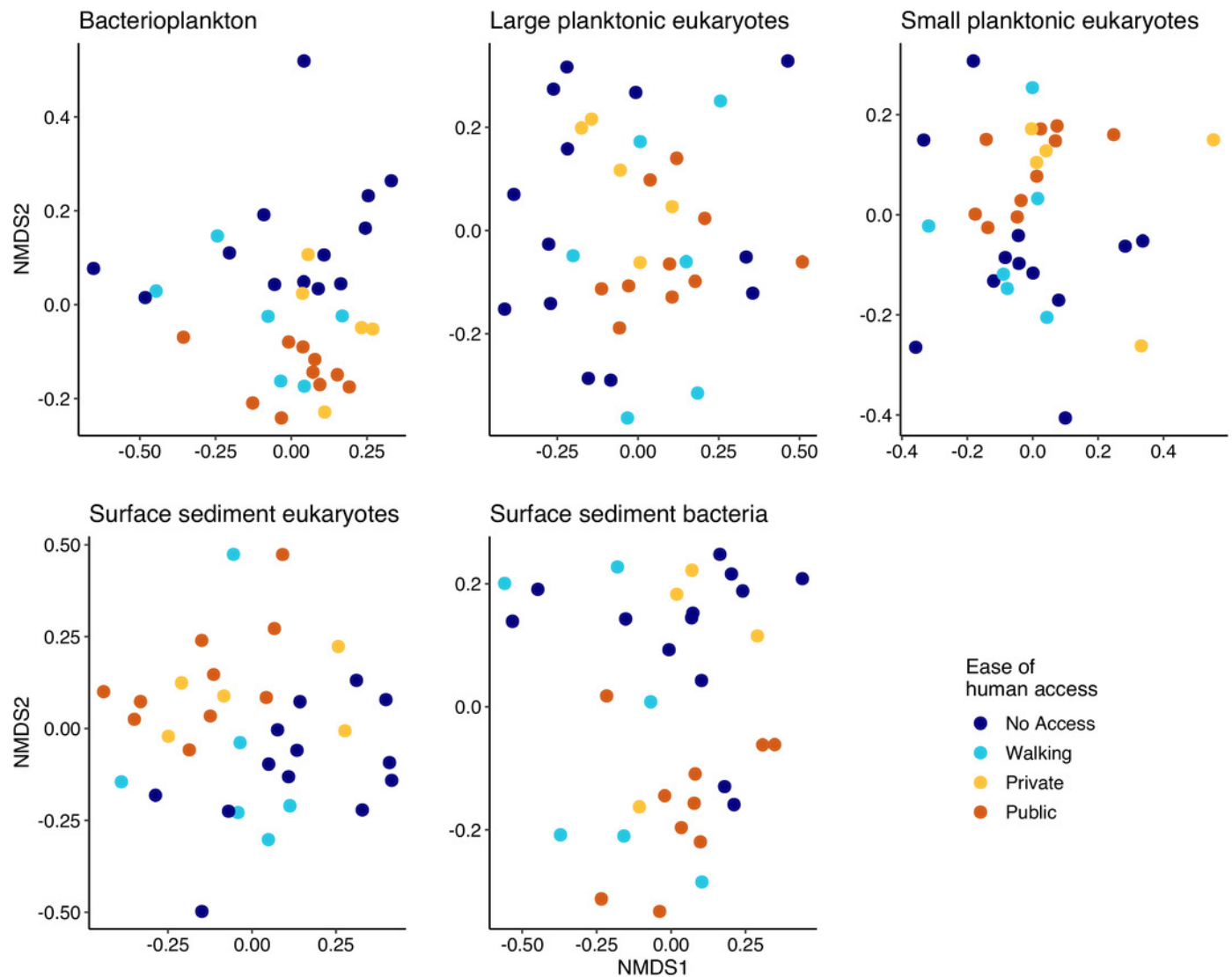



Figure 5

Non metric multidimensional scaling (NMDS) plots for the five biological components.

Points are coloured by the ease of human access to the lake.

The proportion of lake pairwise comparisons attributed to each assembly process for the different biological components assessed in the study.

VS = variable selection HS = homogeneous selection, DL = dispersal limitation, HD = homogenizing dispersal.

	Deterministic		Stochastic		No Dominant Process
	VS (%)	HS (%)	DL (%)	HD (%)	(%)
Bacterioplankton	46.9	13.5	12.1	7.1	20.3
Large planktonic eukaryotes	25.8	30.3	5.5	0.9	37.4
Small planktonic eukaryotes	9.7	18	11	7.2	54.2
Sediment Eukaryotes	40.1	10.2	17.1	2.5	30.1
Sediment Bacteria	73.3	8.1	6.4	8	4.2

The relative abundance and standard deviation of clade III of SAR11 and Daphnia pulex depending on the ease of human access to the lakes.

Ease of human access	Clade III of SAR11		*Daphnia pulex*	
	Mean relative abundance	Standard deviation	Mean relative abundance	Standard deviation
No Access	0	0	0.17	0.62
Walking	4.8	8.31	27.7	22.7
Private	6.43	13.6	0.003	0.003
Public	9.72	10.3	29.2	44