Bacterial community composition and dynamics spanning five years in freshwater bog lakes

Alexandra M. Linz[1] #, Benjamin C. Crary[1], Ashley Shade[2], Sarah Owens[3], Jack A. Gilbert[3.4.5],

Rob Knight[6,7,8], and Katherine D. McMahon[1.9]

[1]Department of Bacteriology, University of Wisconsin – Madison, Madison, WI, 53706, USA

[2]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, 48824, USA

[3]Biosciences Division, Argonne National Laboratory, Argonne, IL 60439, USA

[4]Computation Institute, University of Chicago, Chicago, IL 60637, USA

[5]Department of Ecology and Evolution, Department of Surgery, University of Chicago, Chicago, IL 606037

[6]Center for Microbiome Innovation, Jacobs School of Engineering, University of California, San Diego, La Jolla, California 92093, USA.

[7]Department of Pediatrics, University of California, San Diego School of Medicine, La Jolla, California 92093, USA.

[8]Department of Computer Science and Engineering, Jacobs School of Engineering, University of California San Diego, La Jolla, California 92093, USA.

[9]Department of Civil and Environmental Engineering, University of Wisconsin – Madison, Madison, WI, 53706

corresponding author: Alexandra Linz, Department of Bacteriology, University of Wisconsin

1550 Linden Drive Room 5525 Madison, WI 53706, Email: amlinz@wisc.edu

Running title: Bacterial communities spanning five years in bog lakes

Keywords:

Abstract word count: 186

Importance word count: 141

Article word count: 4795

Abstract

Bacteria play a key role in freshwater biogeochemical cycling, but long-term trends in freshwater bacterial community composition and dynamics are not yet well characterized. We used a multi-year time series of 16S rRNA gene amplicon sequencing data from eight bog lakes to census the freshwater bacterial community and observe annual and seasonal trends in abundance. Multiple sites and sampling events were necessary to begin to fully describe the bacterial communities. Each lake and layer contained a distinct bacterial community, with distinct levels of richness and indicator taxa that likely reflected the environmental conditions of each site. The community present in each year and site was also unique. Despite high interannual variability in community composition, we detected a core community of ubiquitous freshwater taxa. Although trends in abundance did not repeat annually, each freshwater lineage within the communities had a consistent lifestyle, defined by persistence, abundance, and variability. The results of our analysis emphasize the importance of long-term observations, as analyzing only a single year of data would not have allowed us to describe the dynamics and composition of these freshwater bacterial communities to the extent presented here.

Importance

Lakes are excellent systems for investigating bacterial community dynamics because they have clear boundaries and strong environmental gradients. The results of our research demonstrate that bacterial community dynamics operate on multi-year timescales, a finding which likely applies to other ecosystems, with implications for study design and interpretation. Understanding the drivers and controls of bacterial communities on long time scales would improve both our knowledge of fundamental properties of bacterial communities, and our ability to predict community states. In this specific ecosystem, bog lakes play a disproportionately large role in

54 global carbon cycling, and the information presented here may ultimately help refine carbon

55 budgets for these lakes. Finally, all data and code in this study are publicly available. We hope

56 that this will serve as a resource to anyone seeking to answer their own microbial ecology

57 questions using a multi-year time series.

58

59 Introduction

60   One of the major goals of microbial ecology is to predict bacterial community composition.

61 However, we have only a cursory knowledge of the factors that would allow us to predict

62 bacterial community dynamics. To characterize the diversity and dynamics of an ecosystem's

63 bacterial community, sampling the same site multiple times is as necessary as sampling replicate

64 ecosystems. Additionally, the sampling frequency must match the rate of change of the process

65 being studied. We must first understand the scales on which bacterial communities change before

66 we can design experiments that capture a full range of natural variation.

67   Bacterial communities have the potential to change more quickly than communities of

68 macro-organisms due to their fast rate of reproduction. A meta-analysis of time series spanning

69 one to three years found positive species-time relationships, indicating that more taxa are

70 observed as the duration of sampling increases, either due to incomplete sampling, extinction and

71 immigration, or speciation (1). Bacterial time series display time decay, meaning that the

72 community continues to become more dissimilar from the initial sampling event as time from

73 that event increases (2). In one freshwater lake, the amount of change in the bacterial community

74 over a single day was equivalent to dissimilarity between sampling points ten meters apart (3).

75 Conversely, bacterial communities can also change gradually over extremely long time scales, as

76 they are sensitive to changes in environmental parameters such as nutrient availability and

77   temperature. Wetland ecosystems and their carbon emissions are expected to change on scales

78   greater than 300 years (4); as these emissions are the result of bacterial processes, we expect that

79   the bacterial community will change on the same time scale as its ecosystem. Changes in marine

80   phytoplankton regimes have been observed to occur over the past millennium, correlating with

81   shifts in climate (5). With such a large range of potential change, we now recognize the need to

82   more rigorously consider the duration and frequency of sampling in microbial ecology.

83   Multi-year studies of bacterial communities are less common due their logistical

84   difficulties and the need for stable funding, but results from the United States National Science

85   Foundation funded Microbial Observatory projects are exemplary. As a few examples among

86   many, the San Pedro North Pacific - Microbial Observatory contributed to our understanding of

87   heterogeneity of bacterial communities across space and time (6), while research at the Sapelo

88   Island – Microbial Observatory has led the field in linking genomic data to metadata (7). In our

89   own North Temperate Lakes – Microbial Observatory, based in Wisconsin, USA, a multi-year

90   time series of metagenomic data was used to study sweeps in diversity at the genome level (8),

91   adding to our knowledge of how genetic mutation influences bacterial communities. Long-term

92   microbial ecology studies have a time-tested role in the quest to forecast bacterial communities.

93   Our North Temperate Lakes - Microbial Observatory time series was collected from eight

94   bog lakes near Minocqua in the boreal region of northern Wisconsin, USA. Bog lakes contain

95   high levels of dissolved organic carbon in the form of humic and fulvic acids, resulting in dark,

96   "tea-colored" water. Due to their dark color, bog lakes absorb heat from sunlight, resulting in

97   strong stratification during the summer. The top layer in a stratified bog lake, called the

98   "epilimnion," is oxygen-rich and warm. At the lake bottom, an anoxic, cold layer called the

99   "hypolimnion" is formed. The transitions between mixing of these two layers and stratification

100   occur rapidly in these systems, and at different frequencies (called mixing regimes) depending on

101   the depth, surface area, and wind exposure of the lake. Changes in bacterial community

102   composition along the vertical gradients established during stratification are well documented (9,

103   10). Mixing has been shown to be a disturbance to the bacterial communities in bog lakes (11).

104   The bacterial communities in bog lakes are still being characterized, but contains both ubiquitous

105   freshwater organisms (12, 13) and members of the candidate phyla radiation (14). Seasonality in

106   these systems has been suggested (15, 16); however, multiple years of sampling are needed to

107   confirm these prior findings.

108   Our dataset is comprised of 1,387 16S rRNA gene amplicon sequencing samples,

109   collected from eight lakes and two thermal layers over five years. Our primary goal for this

110   dataset was to census the bog lake community and determine which taxa are core to all bog

111   lakes, to each thermal layer, and to each mixing regime. We also sought to learn how mixing

112   regime structures the bacterial community, with our specific hypothesis being that lakes with

113   intermediate levels of disturbance via mixing would be the most diverse. Finally, we investigated

114   seasonality both at the community level and in individual taxa to identify annual trends. This

115   extensive, long-term sampling effort establishes a time series that allows us to assess variability,

116   responses to disturbance and re-occurring trends in freshwater bacterial communities.

117

Results

Overview of community composition

120   A time series of 16S amplicon data recovered from 1,387 samples was used to investigate

121   bacterial community composition over time and across lakes. A total of 8,795 OTUs were

122   detected. As is typical for most freshwater ecosystems, Proteobacteria, Actinobacteria,

123   Bacteroidetes, and Verrucomicrobia were the most abundant phyla (Figure S1). Within these

124   phyla, OTU abundance was highly uneven. For example, much of the abundance of

125   *Proteobacteria* could be attributed to OTUs belonging to the well-known freshwater groups

126   *Polynucleobacter* and *Limnohabitans,* and the freshwater clade acI contributed

127   disproportionately to the observed abundance of Actinobacteria.   Like many microbial

128   communities, unevenness was a recurring theme in this dataset, which had a long rare tail of

129   OTUs and trends driven largely by the most abundant OTUs (17, 18). Trimming of rare taxa did

130   not impact the clustering observed in ordinations, such as those present in Figure 2, even when

131   taxa observed less than 1000 times were removed.

Community richness

133   We hypothesized that disturbance frequency, indicated by mixing regime, determines

134   biodiversity levels. Observed richness was calculated for every sample at the OTU level, and

135   samples were aggregated by lake and layer. Hypolimnia were generally richer than epilimnia

136   (Figure 1, Table S1). Significant differences in richness between lakes were detected. For both

137   layers, polymictic lakes had the fewest taxa, meromictic lakes had the most taxa, and dimictic

138   lakes had intermediate numbers of taxa. This dataset includes two fall mixing events (Trout Bog

139   2007 and North Sparkling Bog 2008), as well as the artificial mixing event in North Sparkling

140   Bog 2008 (11). Richness decreased sharply in mixed samples compared to those taken during the

141   summer stratified period (Figure S2).

Clusters of community composition

143   When differences in community composition were quantified using weighted UniFrac

144   distance and visualized using principal coordinates analysis, several trends emerged. The large

145   number of samples precluded much interpretation using a single PCoA, but sample clustering by

146   layer, mixing regime, and lake was evident (Figure S3). Thus, we also examined PCoA for single

6

147   lakes (both layers). Communities from the epilimnion and hypolimnion layers were significantly

148   distinct from each other at p < 0.05 in all lakes except for polymictic Forestry Bog (FB) (p =

149   0.10) (Figure S4).

150   Within layers, mixing regime was the next factor explaining differences in community

151   composition (Figure 2). Clustering by mixing regime was significant by PERMANOVA in both

152   epilimnia and hypolimnia samples (r2 = 0.20 and r2 = 0.22, respectively, and p = 0.001 in both

153   groups). Lake was a strong factor explaining community composition, with significant cluster in

154   epilimnia (p = 0.001, r2 = 0.34) and hypolimnia (p = 0.001, r2 = 0.49).

155   **Variability and dispersion**

156   While community composition was distinct by layer, lake, and mixing regime, there was

157   still variability in community composition over time. Each year in each lake had a significantly

158   different community composition, indicating interannual variability in the community

159   composition (Figure 3a-c, Figure S5). We found no evidence of repeating seasonal trends during

160   the stratified summer months in these lakes. Likewise, we examined the abundance trends of the

161   most abundant individual OTUs and they did not seem to repeat each year, even when

162   abundances in each year were normalized using z-scores (Figure S6).

163   Varibility can also be assessed by measuring the dispersion of groups in PCoA.

164   Dispersion is the distance of each point from the centroid of a group on an ordination plot. This

165   analysis showed that layers had significantly different degrees of dispersion in two of the

166   dimictic lakes (Trout Bog and South Sparkling Bog) and a meromictic lake (Mary Lake) (Figure

167   3d). Two outliers in Mary Lake were removed; these dates showed different community

168   compositions dominated by few taxa, possibily the result of a bloom event. Dispersion was not

169   significantly different in the polymictic lakes, dimictic North Sparkling Bog, and meromictic

170   Hell's Kitchen. Increased sampling may reveal significant dispersion in these lakes. When

171   dispersion between layers was significant, the epilimnion was on average more dispersed than

172   the hypolimnion, indicating higher variability. This is consistent with our previously published

173   results, and confirms that epilimnia are more variable than hypolimnia.

174   **The core community of bog lakes**

175   One of the goals of this study was to determine the core bacterial community of bog lakes

176   in general, and to determine if mixing regime affects core community membership. Our previous

177   analyses showed that community composition was distinct in each layer and lake (Figure 2),

178   while variability was observed within the same lake and layer (Figure 3). This prompted us to

179   ask whether we had adequately sampled through time and space to fully census the lakes. Still,

180   rarefaction curves generated for the entire dataset and for each layer begin to level off,

181   suggesting that we have indeed sampled the majority of taxa found in our study sites (Figure S7).

182   To identify the taxa that comprise the bog lake core community, we defined "core" as being

183   present in 90% of a group of samples, regardless of abundance in the fully curated dataset. Four

184   OTUs met this criteria for all samples in the dataset: OTU0076 (bacI-A1), OTU0097 (PnecC),

185   OTU0813 (acI-B2), and OTU0678 (LD28). These taxa were therefore also core to both epilimnia

186   and hypolimnia. Additional taxa core to epilimnia also included OTU0004 (betI), OTU0184

187   (acI-B3), OTU0472 (Lhab-A4), and OTU0522 (alfI-A1), while additional hypolimnia core taxa

188   included OTU0042 (Rhodo), OTU0053 (unclassified Verrucomicrobia), and OTU0189 (acI-B2).

189   We performed the same core analysis after combining OTUs assigned to the same tribe

190   (defined by 97% nucleotide similarity in the full length 16S region and phylogenetic branch

191   structure (19)) into new groups. This revealed that certain tribes were core to the entire dataset or

192   thermal layer even though their member OTUs were specific to certain sites. Notably, some

193   OTUs were endemic to specific lakes, even though their corresponding tribe was found in

194   multiple lakes/layers. OTUs not classified at the tribe level were not included. Results were

8

195   similar to those observed at the OTU level, but yielded more core taxa. Tribes core to all samples

196   included bacI-A1, PnecC, acI-B2, and LD28, but also betIII-A1 and acI-B4. In epilimnia, the

197   core tribes were bacI-A1, PnecC, betIII-A1, acI-B3, acI-B2, Lhab-A4, alfI-A1, LD28, and acI-

198   B4, while in hypolimnia, they were Rhodo, bacI-A1, PnecC, betIII-A1, acI-B2, and acI-B4.

199   These results show that despite lake-to-lake differences and interannual variability, there are

200   bacterial taxa that are consistently present in bog lakes.

201   Principal coordinates analysis suggested that samples clustered also by mixing regime

202   (Figure 2). We thus evaluated Venn diagrams of OTUs shared by and unique to each mixing

203   regime to better visualize the overlap in community composition (Figure 4). In both epilimnia

204   and hypolimnia, meromictic lakes had the greatest numbers of unique OTUs while polymictic

205   lakes had the least, consistent with the differences in richness between lakes (Figure 1). Shared

206   community membership, i.e. the number of OTUs present at any abundance in both

207   communities, differed between mixing regimes. Epilimnia (A) and hypolimnia (B) showed

208   similar trends in shared membership: meromictic and dimictic lakes shared the most OTUs,

209   while meromictic and polymictic lakes shared the least.

210   We next used indicator analysis to identify the taxa unique to each mixing regime.

211   Indicator analysis is a statistical method used to determine if taxa are found significantly more

212   frequently in certain pre-determined groups of samples than in others. In this case, the groups

213   were defined by mixing regime, and normalization was applied to account for different numbers

214   of samples in each group. OTUs were grouped at every taxonomic level, and all taxonomic

215   levels were run in the indictor analysis at once to account for differences in the ability of these

216   levels to serve as indicators (for example, the order Actinomycetales is a stronger indicator of

217   polymictic lakes than the phylum Actinobacteria). An abundance threshold of 500 reads was

218   imposed on each taxonomic group. The full table of results from the indicator analysis is

219   available in the supplemental material, while a few indicator taxa of interest are highlighted here.

220   The lineage acI is a ubiquitous freshwater group, with specific clades and tribes showing

221   a preference for bog lakes in previous studies (20, 21). Our dataset shows a further distinction of

222   acI by mixing regime in epilimnia; acI-A tribes were found predominantly in meromictic lakes,

223   with exception of Phila, which is an indicator of polymictic lakes. Tribes of acI-B, particularly

224   OTUs belonging to acI-B2, were indicators of dimictic lakes. Methylophilales, a putative

225   methylotroph, was also an indicator of dimictic lakes, as was putative sulfate reducer

226   Desulfobulbaceae. The phyla Planctomyces, Omnitrophica (formerly OP3), OP8, and

227   Verrucomicrobia were found more often in meromictic lakes, as were putative sulfate reducers

228   Syntrophobacterales and Desulfobacteraceae. Indicators of polymictic lakes include ubiquitous

229   freshwater groups such as Limnohabitans, Polynucleobacter (PnecC), betI-A, and verI-A. These

230   indicator taxa likely reflect the environmental conditions unique to each mixing regime.

231   **Lifestyles of freshwater lineages**

232   Even though OTUs do not show the same trends each year, they do possess patterns that

233   are consistent between years and lakes. We quantified mean abundance when present,

234   persistence (defined as the proportion of samples containing the group of interest), and the

235   coefficient of variation for lineages classified using the freshwater taxonomy, metrics which

236   have been previously used to categorize OTUs (22, 23). Using only these well-defined

237   freshwater groups allowed better taxonomic resolution as we summed the abundances of OTUs

238   by their lineage classification. Lifestyle traits of lineages were consistent across both lakes and

239   years. Low persistence was associated with high variability, and low variability was associated

240   with high abundance (Figure 5, Figure S8). We rarely observed "bloomers," situations where a

241   clade had both high abundance and low persistence; one potential reason for this could be that

10

242   true "bloomers" drop below the detection limit of our sequencing methods when not abundant.

243   Most freshwater lineages were highly persistent at low abundances with low variability. Lineage

244   gamIII of the Gammaproteobacteria was an exception, with low persistence, low abundance, and

245   high variability. Lineages gamI and verI-A occasionally also exhibited this profile. Lineages

246   betII and acI were highly abundant and persistent with low variability, consistent with their

247   suggested lifestyles as ubiquitous freshwater generalists (12, 21).

248

Discussion

250   The North Temperate Lakes - Microbial Observatory dataset is a comprehensive 16S

251   amplicon survey spanning four years, eight lakes, and two thermal layers. We found that

252   multiple years of sampling were necessary to census the community of bog lake ecosystems.

253   Richness and membership in these communities were structured by layer, mixing regime, and

254   lake. We identified specific bacterial taxa present throughout the dataset, as well as taxa endemic

255   to certain depths or mixing regimes. Mixing events were associated with reduced richness and an

256   increase in the proportion of certain taxa. High levels of variability were detected in this dataset;

257   each year in each lake harbored a unique bacterial community. However, freshwater lineages

258   still showed consistent lifestyles, defined by abundance, persistence, and variability, across lakes

259   and years, even though the abundance trends of individual populations were . Our results

260   emphasize the importance of multiple sampling events to assess full bacterial community

261   membership and variability in an ecosystem.

262   The bog lakes in this study have been model systems for freshwater microbial ecology for

263   many years. Early studies used Automated Ribosomal Intergenic Spacer Analysis (ARISA), a

264   fingerprinting technique for identifying unique bacterial taxa in environmental samples (24). Our

265 research built upon these studies and added information about the taxonomic identities of

266 bacterial groups. For example, persistent and unique bacterial groups were detected in the bog

267 lakes using ARISA (25); using 16S amplicon sequencing, we also found persistent groups and

268 could identify them as the ubiquitous freshwater bacteria LD28, acI-B2, PnecC, and bacI-A1.

269 Differences in richness and community membership were previously detected between Crystal

270 Bog, Trout Bog, and Mary Lake, three sites representative of the three mixing regime categories

271 of polymictic, dimictic, and meromictic (25). Our data supported these results and suggest that

272 these trends are indeed linked with mixing regime, as we included multiple lakes of each type

273 sampled over multiple years in this study.

274   We also supported previous research on the characteristics of bacterial communities in

275 the epilimnion and hypolimnion, and the impacts of lake mixing on these communities. We

276 confirmed that epilimnia communities tended to be more dispersed than hypolimnia

277 communities, potentially due to increased exposure to climatic events (25). Mixing was

278 disruptive to both epilimnion and hypolimnion communities, selecting for only a few taxa that

279 persist during this disturbance, but quickly recovering diversity once stratification was re-

280 established (11, 26). Comparing richness between lakes of different mixing regimes did not

281 support the intermediate disturbance hypothesis, which was our initial inspiration for the

282 collection of this dataset; rather, the least frequently mixing lakes had the most diverse

283 communities. As many variables co-vary with mixing regime (such as depth, volume of

284 integrated water column, dissolved carbon concentrations and total nitrogen concentration), it is

285 not clear which variables are driving this trend. One likely explanation is that increased depth

286 leads to increased habitat heterogeneity, as more distinct niches develop along the vertical

287 chemical gradients in the lake. These results are in line with current thinking in the ecological

12

288 community, as other studies on diversity and disturbance have also found no evidence for the

289 intermediate disturbance hypothesis (27).

290   We were not able to detect repeatable annual trends in bog lakes in our multiple years of

291 sampling. While seasonality in marine and river systems has been well-established by our

292 colleagues, previous research on seasonality in freshwater lakes has produced inconsistent results

293 (28–31). Distinct, seasonally repeatable community types were identified in alpine lakes, but

294 stratified summer communities were distinct each year (32). Seasonal trends were detected in a

295 time series from Lake Mendota similar to this study, but summer samples in Lake Mendota were

296 more variable then those collected in other seasons (33). In the previous ARISA-based research

297 on the bog lakes in our dataset, community properties such as richness and rate of change were

298 consistent each year, and the phytoplankton communities were hypothesized to drive seasonal

299 trends in the bacterial communities based on correlation studies (34–36). Synchrony in seasonal

300 trends was observed (35); however, in a second year of sampling for seasonal trends in Crystal

301 Bog and Trout Bog, these findings were not reproduced (37). Successional trends were studied in

302 Crystal Bog and Lake Mendota with a relatively small number of samples collected over two

303 years and "dramatic changes" in community composition associated with drops in biodiversity

304 were described during the summer months, while spring, winter, and fall had more stable

305 community composition (34). Because our dataset was sparsely represented by seasons other

306 than summer, higher summer variability may explain why we see a different community each

307 year and a lack of seasonal trends in community composition. However, we cannot disprove the

308 influence of seasonality on bacterial community dynamics in temperate freshwater lakes.

309   One of the biggest benefits of 16S rRNA gene amplicon sequencing over ARISA is the

310 ability to assign names to sequences. In addition to a core of persistent taxa found in nearly every

13

311   sample collected, we also identified taxa endemic to either the epilimnion or hypolimnion and to

312   specific mixing regimes. These endemic taxa likely reflect the biogeochemical differences driven

313   by mixing regime. Dimictic and meromictic hypolimnia, which are consistently anaerobic,

314   harbor putative sulfur and sulfate reducers not present in polymictic hypolimnia, which are more

315   frequently oxygenated. Members of the acI lineage partition by mixing regime in epilimnia, and

316   the functional traits driving this filtering effect are the subject of active study (20). Interestingly,

317   the meromictic Mary Lake hypolimnion contains several taxa classified into the candidate phyla

318   radiation and a larger proportion of completely unclassified reads than other hypolimnia (38).

319   This is consistent with the findings of other 16S and metagenomics studies of meromictic lakes

320   (39, 40), and suggests that the highly reduced and consistently anaerobic conditions in

321   meromictic hypolimnia are excellent study systems for research on members of the candidate

322   phyla radiation and "microbial dark matter".

323   Perhaps the biggest implication of this study is the importance of repeated sampling of

324   microbial ecosystems. A similar dataset spanning only a single year would not have captured the

325   full extent of variability observed, and therefore would not have detected as many of the taxa

326   belonging to the bog lake community; even our four years of weekly sampling did not result in

327   level rarefaction curves (Figure S7). While we found no evidence for seasonal trends or repeated

328   annual trends, it is possible that there are cycles or variables acting on scales longer than the five

329   years covered in this dataset, or that annual differences are driven by environmental factors that

330   do not occur every year. Unmeasured biotic interactions between bacterial taxa may also

331   contribute to the observed variability. Understanding the factors that contribute to variability in

332   bog lake communities will lead to improved predictive modelling in freshwater systems,

333   allowing forecasting of bloom events and guiding better management strategies. Additionally,

14

334   these systems may be ideal for addressing some of the core questions in microbial ecology, such

335   as how community assembly occurs, how interactions between taxa shape community

336   composition, and how resource partitioning drives the lifestyles of bacterial taxa.

337   To answer these questions and more, we continue to collect and sequence samples for the

338   North Temperate Lakes – Microbial Observatory, and we are expanding our sequencing

339   repertoire beyond 16S rRNA gene sequencing. All 16S rRNA gene data we have currently

340   generated can be found in the R package "OTUtable" which is available on CRAN for

341   installation via the R command line, or on our GitHub page. This dataset has already been used

342   in a meta-analysis of microbial time series (1). We hope that this dataset and its future expansion

343   will be used as a resource for researchers investigating their own questions about how bacterial

344   communities behave on long time scales.

345

Materials and Methods

346

347   **Sample Collection**

348   Water was collected from eight bog lakes during the summers of 2005, 2007, 2008 and

349   2009, as previously described (25). Briefly, the epilimnion and hypolimnion layers were

350   collected separately using an integrated water column sampler. Dissolved oxygen and

351   temperature profiles were measured at the time of collection using a handheld YSI 550A (YSI

352   Inc., Yellow Springs, OH). After transport to the laboratory, approximately 150 mL from each

353   well-mixed sample was filtered through a 0.22 micron polyethersulfone filter (Supor 200, Pall,

354   Port Washington, NY). Filters were stored at -80C until DNA extraction using FastDNA Spin Kit

355   for Soil (MP Biomedicals, Santa Ana, CA), with minor modifications (41). The sampling sites

356   are located near Boulder Junction, WI, and were chosen to include lakes represent the three

15

357 mixing regimes of polymictic (multiple mixing events per year), dimictic (two mixing events per

358 year, usually in spring and fall), and meromictic (no record mixing events) (Table 1). Trout Bog

359 and Crystal Bog are also primary study sites for the North Temperate Lakes - Long Term

360 Ecological Research Program, which measures a suite of chemical limnology parameters

361 fortnightly during the open water season. The NTL-LTER also maintains autonomous sensing

362 buoys on Trout Bog and Crystal Bog, allowing for more refined mixing event detection based on

363 thermistor chain measurements.

364 **Sequencing**

365 A total of 1,510 DNA samples, including 547 biological replicates, were sequenced by

366 the Earth Microbiome Project according to their standard protocols in 2010, using the original

367 V4 primers (FWD:GTGCCAGCMGCCGCGGTAA;REV:GGACTACHVGGGTWTCTAAT)

368 (42). Briefly, the V4 region was amplified and sequenced using Illumina HiSeq, resulting in

369 77,517,398 total sequences with an average length of 150 base pairs. To reduce the number of

370 erroneous sequences, QIIME's "deblurring" algorithm for reducing sequence error in Illumina

371 data was applied (43). Based on the sequencing error profile, this algorithm removes reads that

372 are likely to be sequencing errors if those reads are both low in abundance and highly similar to a

373 high abundance read. Reads occurring less than 25 times in the entire dataset were removed after

374 deblurring, leaving 9,856 unique sequences. These sequences are considered operational

375 taxonomic units (OTUs).

376 570 sequences with long homopolymer runs, ambiguous base calls, or incorrect sequence

377 lengths were found and removed via mothur v1.34.3 (44). Thirty-three chimeras and 340

378 chloroplast sequences (based on pre-clustering and classification with the Greengenes 16S

379 database, May 2013) (45) were removed. Samples were rarefied to 2,500 reads; samples with

380 less than 2,500 reads were omitted, resulting in 1,387 remaining samples. The rarefaction cutoff

16

381   used was determined based on the results of simulation; 2,500 reads was chosen to maximize the

382   number of samples retained, while maintaining sufficient quality for downstream analysis of

383   diversity metrics.

384   Representative sequences for each OTU were classified in either our curated freshwater

385   database (19) or the Greengenes database based on the output of NCBI-BLAST (blast+ 2.2.3.1)

386   (46). Representative sequences from each OTU were randomly chosen. The program blastn was

387   used to compare representative sequences to full-length sequences in the freshwater database.

388   OTUs matching the freshwater database with a percent identity greater than 98% were classified

389   in that database, and remaining sequences were classified in the Greengenes database. Both

390   classification steps were performed in mothur using the Wang method (47), and classifications

391   with less than 70% confidence were not included. A detailed workflow for quality control and

392   classification of our sequences is available at (https://github.com/McMahonLab/16STaxAss)

393   (manuscript in prep).

394   **Statistics**

395   Statistical analysis was performed in R v3.3.2 (R Development Core Team, 2008. R: A

396   language and environment for statistical computing.). Significant differences in richness between

397   lakes was tested using a pairwise Wilcoxon sum rank test with a Bonferroni adjustment in the R

398   package "exactRankTests" (T. Hothorn and K. Hornik, 2015. exactRankTests: Exact

399   Distributions for Rank and Permutation Tests). Similarity between samples was compared using

400   UniFrac distances, as implement in "phyloseq" (48) (P.J. McMurdie and S. Holmes, 2013.

401   phyloseq: An R Package for reproducible interactive analysis and graphic of microbiome census

402   data). Weighted and unweighted Unifrac distance (48) was compared with Bray-Curtis

403   Dissimilarity and Jaccard Similarity, implemented in "vegan" (J. Oksanen, 2016. vegan:

404   Community Ecology Package). Weighted UniFrac distances were chosen for principle

17

405   coordinates analysis, performed by betadisper() in "vegan", because it explained the greatest

406   amount of variation in the first two axes.   Significant clustering by year in PCoA and in

407   dispersion between lakes was tested using PERMADISP with the function adonis() in "vegan."

408   Indicator species analysis was performed using "indicspecies" (49). Only taxa with read

409   abundances of at least 500 reads in the entire dataset were used for this analysis. The group-

410   normalized coefficient of correlation was chosen for this analysis because it measures both

411   positive and negative habitat preferences and accounts for differences in the number of samples

412   from each site. All taxonomic levels were included in this analysis to determine which level of

413   resolution was the best indicator for each taxonomic group.

414   Plots were generated using "ggplot2" (H. Wickham, 2009. ggplot2: Elegant Graphics for

415   Data Analysis) and "cowplot" (C. Wilke, 2016. cowplot: Streamlined Plot Themes and Plot

416   Annotations for 'ggplot2'). "reshape2" was used for data formatting (H. Wickham, 2007.

417   Reshaping Data with the reshape Package). Data and code from this study can be downloaded

418   from   the   R   package   "OTUtable"   and   the   McMahon   Lab   GitHub   repository

419   "North_Temperate_Lakes-Microbial_Observatory."

420   Acknowledgments

References

438   1.   **Shade A**, **Caporaso JG**, **Handelsman J**, **Knight R**, **Fierer N**. 2013. A meta-analysis of

439   changes in bacterial and archaeal communities with time. ISME J **7**:1493–506.

440   2.   **Faust K**, **Lahti L**, **Gonze D**, **de Vos WM**, **Raes J**. 2015. Metagenomics meets time series

441   analysis: unraveling microbial community dynamics. Curr Opin Microbiol **25**:56–66.

442   3.   **Jones SE**, **Cadkin TA**, **Newton RJ**, **McMahon KD**. 2012. Spatial and temporal scales of

443   aquatic bacterial beta diversity. Front Microbiol **3**:1–10.

444   4.   **Mitsch WJ**, **Bernal B**, **Nahlik AM**, **Mander Ü**, **Zhang L**, **Anderson CJ**, **Jørgensen SE**,

445   **Brix H**. 2013. Wetlands, carbon, and climate change. Landsc Ecol **28**:583–597.

446   5.   **McMahon KW**, **McCarthy MD**, **Sherwood OA**, **Larsen T**, **Guilderson TP**. 2015.

447   Millennial-scale plankton regime shifts in the subtropical North Pacific Ocean. Science

448   (80-) **350**:1530–1533.

449   6.   **Hewson I**, **Steele JA**, **Capone DG**, **Fuhrman JA**. 2006. Remarkable heterogeneity in

450   meso- and bathypelagic bacterioplankton assemblage composition. Limnol Oceanogr

451   **51**:1274–1283.

452   7.   **Gifford SM**, **Sharma S**, **Moran MA**. 2014. Linking activity and function to ecosystem

453   dynamics in a coastal bacterioplankton community. Front Microbiol **5**:1–12.

454   8.   **Bendall ML**, **Stevens SLR**, **Chan L**, **Malfatti S**, **Schwientek P**, **Tremblay J**,

455   **Schackwitz W**, **Martin J**, **Pati A**, **Bushnell B**, **Froula J**, **Kang D**, **Tringe SG**,

456   **Bertilsson S**, **Moran MA**, **Shade A**, **Newton RJ**, **McMahon KD**, **Malmstrom RR**.

457   2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial

458   populations. ISME J **10**:1589–1601.

459   9.   **Taipale S**, **Jones R**, **Tiirola M**. 2009. Vertical diversity of bacteria in an oxygen-stratified

460   humic lake, evaluated using DNA and phospholipid analyses. Aquat Microb Ecol **55**:1–

461   16.

462   10.   **Garcia SL**, **Salka I**, **Grossart HP**, **Warnecke F**. 2013. Depth-discrete profiles of

463   bacterial communities reveal pronounced spatio-temporal dynamics related to lake

464   stratification. Environ Microbiol Rep **5**:549–555.

465   11.   **Shade A**, **Read JS**, **Youngblut ND**, **Fierer N**, **Knight R**, **Kratz TK**, **Lottig NR**, **Roden**

466   **EE**, **Stanley EH**, **Stombaugh J**, **Whitaker RJ**, **Wu CH**, **McMahon KD**. 2012. Lake

467   microbial communities are resilient after a whole-ecosystem disturbance. ISME J **6**:2153–

468   2167.

469   12.   **Hahn MW**, **Scheuerl T**, **Jezberová J**, **Koll U**, **Jezbera J**, **Šimek K**, **Vannini C**, **Petroni**

470   **G**, **Wu QL**. 2012. The passive yet successful way of planktonic life: Genomic and

471   experimental analysis of the ecology of a free-living polynucleobacter population. PLoS

472   One **7**:1–17.

473   13.   **Garcia SL**, **McMahon KD**, **Grossart HP**, **Warnecke F**. 2013. Successful enrichment of

20

474   the ubiquitous freshwater acI Actinobacteria. Environ Microbiol Rep 1–7.

475   14. **Peura S**, **Eiler A**, **Bertilsson S**, **Nyka H**, **Tiirola M**, **Jones RI**. 2012. Distinct and diverse

476   anaerobic bacterial communities in boreal lakes dominated by candidate division OD1.

477   ISME J J **6**:1640–1652.

478   15. **Eiler A**, **Heinrich F**, **Bertilsson S**. 2012. Coherent dynamics and association networks

479   among lake bacterioplankton taxa. ISME J **6**:330–42.

480   16. **Graham JM**, **Kent AD**, **Lauster GH**, **Yannarell AC**, **Graham LE**, **Triplett EW**. 2004.

481   Seasonal dynamics of phytoplankton and planktonic protozoan communities in a northern

482   temperate humic lake: Diversity in a dinoflagellate dominated system. Microb Ecol

483   **48**:528–540.

484   17. **Mariadassou M**, **Pichon S**, **Ebert D**. 2015. Microbial ecosystems are dominated by

485   specialist taxa. Ecol Lett **18**:974–982.

486   18. **Newton RJ**, **Shade A**. 2016. Lifestyles of rarity : understanding heterotrophic strategies to

487   inform the ecology of the microbial rare biosphere. Aquat Microb Ecol **78**:51–63.

488   19. **Newton RJ**, **Jones SE**, **Eiler A**, **McMahon KD**, **Bertilsson S**. 2011. A guide to the

489   natural history of freshwater lake bacteria. Microbiol Mol Biol Rev **75**:14–49.

490   20. **Garcia SL**, **Buck M**, **McMahon KD**, **Grossart H-P**, **Eiler A**, **Warnecke F**. 2015.

491   Auxotrophy and intra-population complementary in the "interactome" of a cultivated

492   freshwater model community. Mol Ecol **24**:4449–4459.

493   21. **Ghylin TW**, **Garcia SL**, **Moya F**, **Oyserman BO**, **Schwientek KT**, **Mutschler J**,

494   **Dwulit-Smith J**, **Chan L-K**, **Martinez-Garcia M**, **Sczyrba A**, **Stepanauskas R**,

495   **Grossart H-P**, **Woyke T**, **Warnecke F**, **Malmstrom R**, **Bertilsson S**, **McMahon KD**.

496   2014. Comparative single-cell genomics reveals potential ecological niches for the

497   freshwater acI Actinobacteria lineage. ISME J **8**:2503–2516.

498   22.   **Herren CM**, **Webert KC**, **McMahon KD**. 2016. Environmental Disturbances Decrease

499   the Variability of Microbial Populations within Periphyton. mSystems **1**:1–14.

500   23.   **Shade AL**, **Gilbert JA**. 2015. Temporal patterns of rarity provide a more complete view

501   of microbial diversity. Trends Microbiol **23**:335–340.

502   24.   **Fisher MM**, **Triplett EW**. 1999. Automated approach for ribosomal intergenic spacer

503   analysis of microbial diversity and its application to freshwater bacterial communities.

504   Appl Environ Microbiol **65**:4630–4636.

505   25.   **Shade A**, **Jones SE**, **McMahon KD**. 2008. The influence of habitat heterogeneity on

506   freshwater bacterial community composition and dynamics. Environ Microbiol **10**:1057–

507   1067.

508   26.   **Shade A**, **Read JS**, **Welkie DG**, **Kratz TK**, **Wu CH**, **McMahon KD**. 2011. Resistance,

509   resilience and recovery: Aquatic bacterial dynamics after water column disturbance.

510   Environ Microbiol **13**:2752–2767.

511   27.   **Fox JW**. 2013. The intermediate disturbance hypothesis should be abandoned. Trends

512   Ecol Evol **28**:86–92.

513   28.   **Crump BC**, **Hobbie JE**. 2005. Synchrony and seasonality in bacterioplankton

514   communities of two temperate rivers. Limnol Oceanogr **50**:1718–1729.

515   29.   **Gilbert JA**, **Steele JA**, **Caporaso JG**, **Steinbruck L**, **Reeder J**, **Temperton B**, **Huse S**,

516   **McHardy AC**, **Knight R**, **Joint I**, **Somerfield P**, **Fuhrman JA**, **Field D**. 2012. Defining

517   seasonal marine microbial community dynamics. ISME J **6**:298–308.

518   30.   **Fuhrman JA**, **Hewson I**, **Schwalbach MS**, **Steele JA**, **Brown M V**, **Naeem S**. 2006.

519   Annually reoccurring bacterial communities are predictable from ocean conditions. Proc

520   Natl Acad Sci USA **103**:13104–13109.

521 31. **Cram J**, **Chow C**, **Sachdeva R**, **Needham D**, **Parada A**, **Steele J**, **Fuhrman J**. 2015.

522   Seasonal and interannual variability of the marine bacterioplankton community

523   throughout the water column over ten years. ISME J **9**:563–580.

524 32. **Nelson CE**. 2009. Phenology of high-elevation pelagic bacteria: the roles of meteorologic

525   variability, catchment inputs and thermal stratification in structuring communities. ISME J

526   **3**:13–30.

527 33. **Kara EL**, **Hanson PC**, **Hu YH**, **Winslow L**, **McMahon KD**. 2013. A decade of seasonal

528   dynamics and co-occurrences within freshwater bacterioplankton communities from

529   eutrophic Lake Mendota, WI, USA. ISME J **7**:680–4.

530 34. **Yannarell AC**, **Kent AD**, **H. LG**, **Kratz TK**, **Triplett EW**. 2003. Temporal Patterns in

531   Bacterial Communities in Three Temperate Lakes of Different Trophic Status. Microb

532   Ecol **46**:391–405.

533 35. **Kent AD**, **Yannarell AC**, **Rusak JA**, **Triplett EW**, **McMahon KD**. 2007. Synchrony in

534   aquatic microbial community dynamics. ISME J **1**:38–47.

535 36. **Kent AD**, **Jones SE**, **Lauster GH**, **Graham JM**, **Newton RJ**, **McMahon KD**. 2006.

536   Experimental manipulations of microbial food web interactions in a humic lake: Shifting

537   biological drivers of bacterial community structure. Environ Microbiol **8**:1448–1459.

538 37. **Rusak JA**, **Jones SE**, **Kent AD**, **Shade A**, **McMahon TD**. 2009. Spatial synchrony in

539   microbial community dynamics : testing among-year and lake patterns. Verh Internat

540   Verein Limnol **30**:936–940.

541 38. **Hug LA**, **Baker BJ**, **Anantharaman K**, **Brown CT**, **Probst AJ**, **Castelle CJ**,

542   **Butterfield CN**, **Hernsdorf AW**, **Amano Y**, **Ise K**, **Suzuki Y**, **Dudek N**, **Relman DA**,

543   **Finstad KM**, **Amundson R**, **Thomas BC**, **Banfield JF**. 2016. A new view of the tree of

544   life. Nat Microbiol **1**:1–6.

545   39.   **Gies EA**, **Konwar KM**, **Beatty JT**, **Hallam SJ**. 2014. Illuminating microbial dark matter

546   in meromictic Sakinaw Lake. Appl Environ Microbiol **80**:6807–6818.

547   40.   **Borrel G**, **Lehours AC**, **Bardot C**, **Bailly X**, **Fonty G**. 2010. Members of candidate

548   divisions OP11, OD1 and SR1 are widespread along the water column of the meromictic

549   Lake Pavin (France). Arch Microbiol **192**:559–567.

550   41.   **Shade A**, **Kent AD**, **Jones SE**, **Newton RJ**, **Triplett EW**, **McMahon KD**. 2007.

551   Interannual dynamics and phenology of bacterial communities in a eutrophic lake. Limnol

552   Oceanogr **52**:487–494.

553   42.   **Caporaso JG**, **Lauber CL**, **Walters WA**, **Berg-Lyons D**, **Huntley J**, **Fierer N**, **Owens**

554   **SM**, **Betley J**, **Fraser L**, **Bauer M**, **Gormley N**, **Gilbert JA**, **Smith G**, **Knight R**. 2012.

555   Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq

556   platforms. ISME J **6**:1621–1624.

557   43.   **Amir A**, **McDonald D**, **Navas-Molina J**, **Kopylova E**, **Morton J**, **Xu Z**, **Kightley E**,

558   **Thompson L**, **Hyde E**, **Gonzalez A**, **Knight R**. 2017. Deblur Rapidly Resolves Single-

559   Nucleotide Community Sequence Patterns. mSystems **2**:1–7.

560   44.   **Schloss PD**, **Westcott SL**, **Ryabin T**, **Hall JR**, **Hartmann M**, **Hollister EB**, **Lesniewski**

561   **RA**, **Oakley BB**, **Parks DH**, **Robinson CJ**, **Sahl JW**, **Stres B**, **Thallinger GG**, **Van**

562   **Horn DJ**, **Weber CF**. 2009. Introducing mothur: Open-source, platform-independent,

563   community-supported software for describing and comparing microbial communities.

564   Appl Environ Microbiol **75**:7537–7541.

565   45.   **DeSantis TZ**, **Hugenholtz P**, **Larsen N**, **Rojas M**, **Brodie EL**, **Keller K**, **Huber T**,

24

566   **Dalevi D**, **Hu P**, **Andersen GL**. 2006. Greengenes, a chimera-checked 16S rRNA gene

567   database and workbench compatible with ARB. Appl Environ Microbiol **72**:5069–5072.

568 46. **Camacho C**, **Coulouris G**, **Avagyan V**, **Ma N**, **Papadopoulos J**, **Bealer K**, **Madden**

569   **TL**. 2009. BLAST plus : architecture and applications. BMC Bioinformatics **10**:1–9.

570 47. **Wang Q**, **Garrity GM**, **Tiedje JM**, **Cole JR**. 2007. Naive Bayesian Classifier for Rapid

571   Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ

572   Microbiol **73**:5261–5267.

573 48. **Lozupone C**, **Knight R**. 2005. UniFrac : a New Phylogenetic Method for Comparing

574   Microbial Communities UniFrac : a New Phylogenetic Method for Comparing Microbial

575   Communities. Appl Environ Microbiol **71**:8228–8235.

576 49. **De Cáceres M**, **Legendre P**. 2009. Associations between species and groups of sites:

577   indices and statistical inference. Ecology **90**:3566–3574.

578

579

580   Tables

581   **Table 1. Location and characteristics of study sites.** The lakes included in this time series are

582   small, humic bog lakes in the boreal region near Minocqua, Wisconsin, USA. They range in

583   depth from 2 to 21.5 meters and encompass a range of water column mixing frequencies (termed

584   regimes). Dimictic lakes mix twice per year, typically in fall and spring, while polymictic lakes

585   can mix more than twice throughout the spring, summer, and fall. Meromictic lakes have no

586   recorded mixing events. pH was measured in 2007, while nutrient data was measured in 2008

587   (with the exceptions of FB, WS, and HK, measured in 2007). When two values are present in a

588   single box, the first represents the epilimnion value and the second represents the hypolimnion

589   value.

590

591

592

	Forestry Bog	Crystal Bog	North Sparkling Bog	West Sparkling Bog	Trout Bog	South Sparkling Bog	Hell's Kitchen	Mary Lake
ID	FB	CB	NS	WS	TB	SS	HK	MA
Depth (m)	2.0	2.5	4.5	4.6	7.0	8.0	19.3	21.5
Surface area (m^2)	1300	5600	4700	11900	10100	4400	30000	12000
Mixing regime	Polymictic	Polymictic	Dimictic	Polymictic	Dimictic	Dimictic	Meromictic	Meromictic
GPS coordinates	46.04777, -89.651248	46.007639, -89.606341	46.004819, -89.705214	46.004633, -89.709082	46.041140, -89.686352	46.041140, -89.709082	46.186674, -89.702510	46.250764, -89.900419
Years sampled	2007	2007, 2009	2007, 2008, 2009	2007	2005, 2007, 2008, 2009	2007, 2008, 2009	2007	2005, 2007, 2008, 2009
pH	4.97, 4.85	4.49, 4.41	4.69, 4.80	5.22, 5.14	4.60, 4.78	4.46, 4.94		5.81, 5.72
Dissolved inorganic carbon (ppm)	0.94, 1.46	0.69, 1.72	1.12, 2.31	0.76, 1.56	1.73, 4.47	1.97, 6.42	2.91, 9.70	5.54, 12.38
Dissolved organic carbon (ppm)	10.22, 8.96	15.47, 13.6	10.05, 10.40	7.26, 7.27	19.87, 20.58	12.40, 21.92	7.26, 7.33	20.63, 67.10
Total nitrogen (ppb)		620.57, 846.00	629.09, 809.45		737.71, 1121.00	813.88, 1498		1332.57, 3652.38
Total phosphorus (ppb)		30.00, 38.86	78.00, 135.45		50.57, 53.25	48.63, 69.14		78.00, 303.50
Total dissolved nitrogen (ppb)		1290.19, 490.13	442.39, 586.56		582.5, 820.21	451.63, 1179.21		1024.5, 3220.14
Total dissolved phosphorus (ppb)		84.25, 14.88	70.22, 22.67		34.5, 31.57	16.25, 18.29		71.13, 228

593

594

27

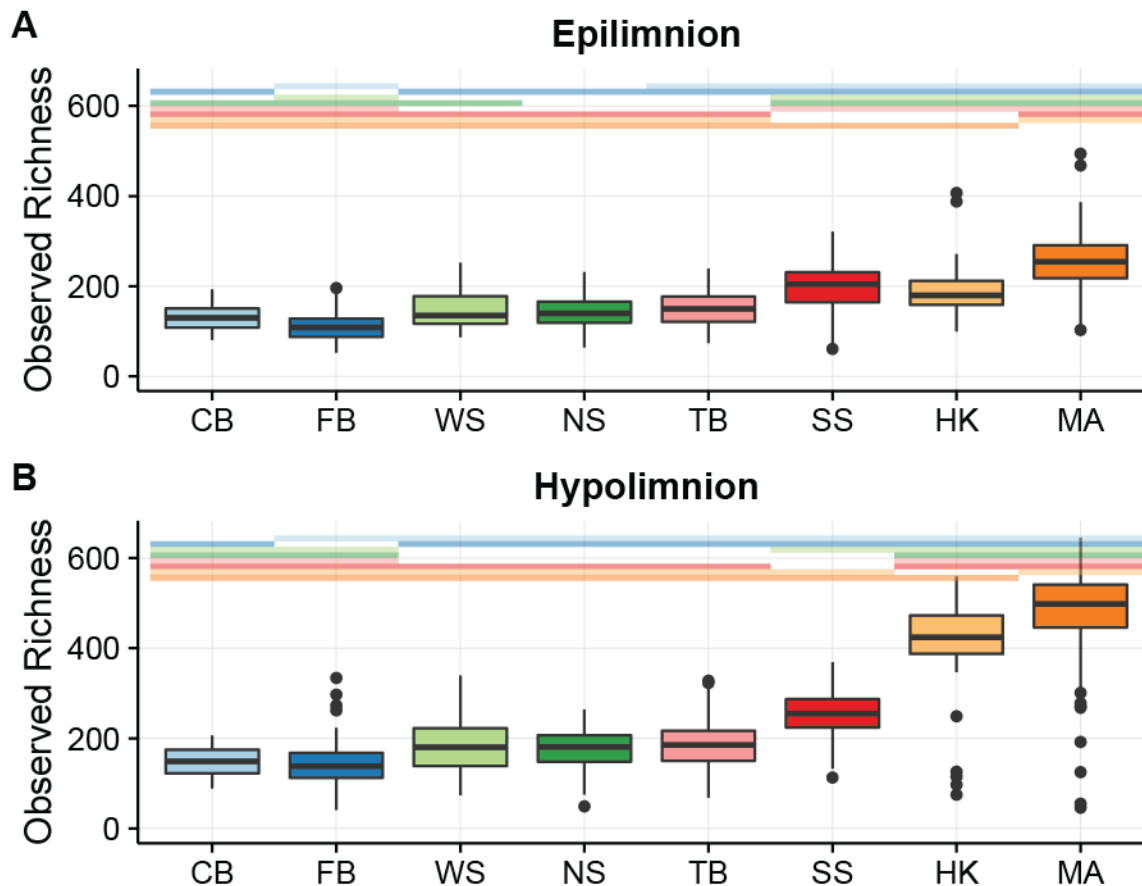595   Figures

596

597

598   **Figure 1. Richness by layer and lake.** Lakes on the x axis are arranged by depth (see Table 1

599   for lake abbreviations and depth measurements). Lakes CB, FB, and WS are polymictic, lakes

600   NS, TB, and SS are dimictic, and lakes HK and MA are meromictic. Colored bars above each

601   plot represent significant differences in richness between lakes, with each colored bar matching

602   the color of a lakes boxplot. For example, in Panel A, the boxplot for CB has the colored bars

603   matching FB, NS, TB, SS, HK, and MA above it. This indicates that it is significantly different

604   from these lakes, but not significantly different from the missing colored bar, WS. The mean and

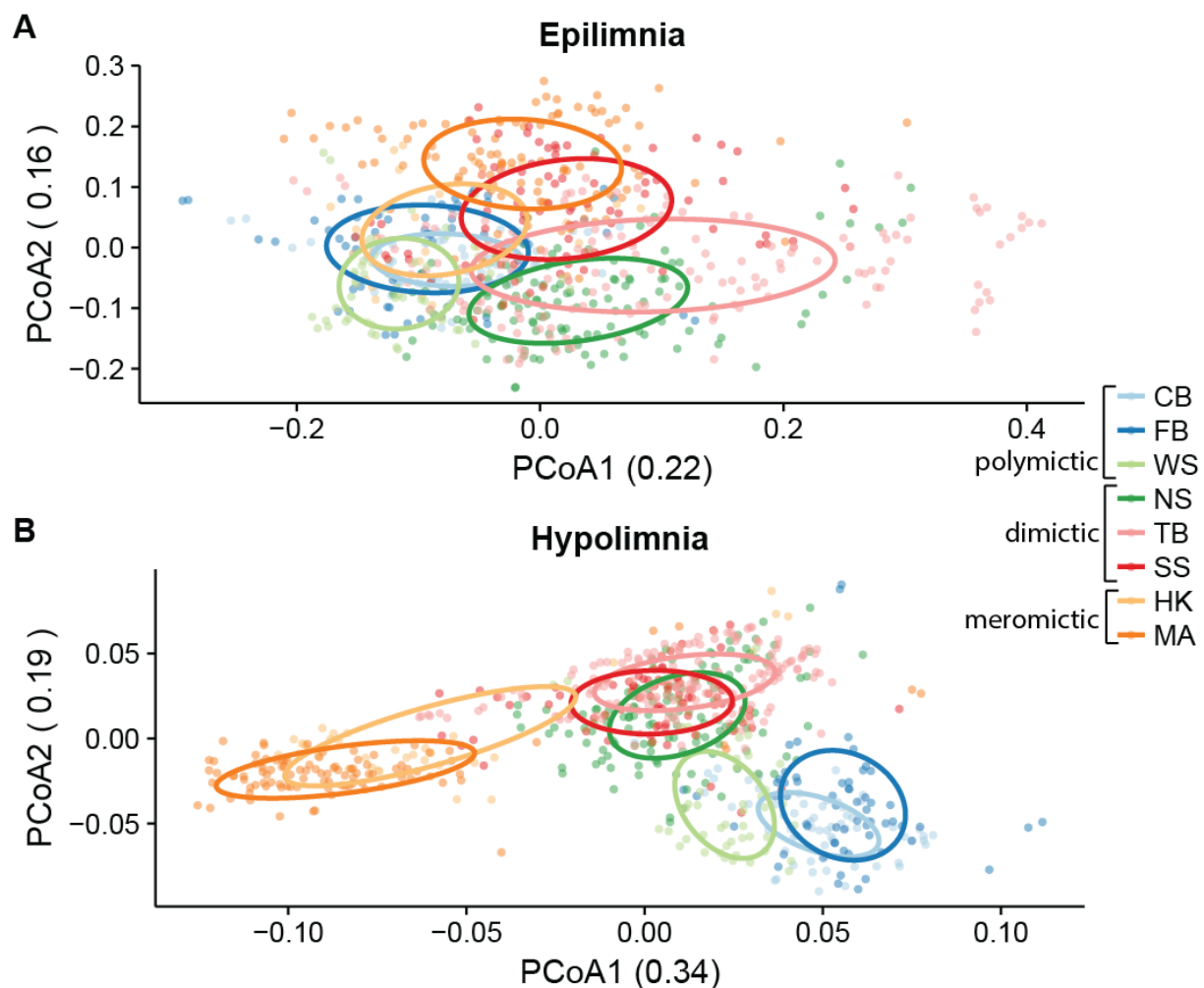605   standard deviation for each lake and layer is reported in Table S1.

28

606

607

608   **Figure 2.** Weighted UniFrac distance was used to perform principal coordinates analysis on

609   epilimnion (A) and hypolimnion (B) samples. In both layers, samples cluster significantly by

610   lake and mixing regime as tested using PERMANOVA. (See Table 1 for lake abbreviations; CB,

611   FB, and WS are polymictic, NS, TB, and SS are dimictic, HK and MA are meromictic). Ellipses

612   indicating the clustering of each lake were calculate based on standard error using a 95%

613   confidence interval. Differences in bacterial community composition between lakes and mixing

614   regimes are more pronounced in hypolimnia than epilimnia.
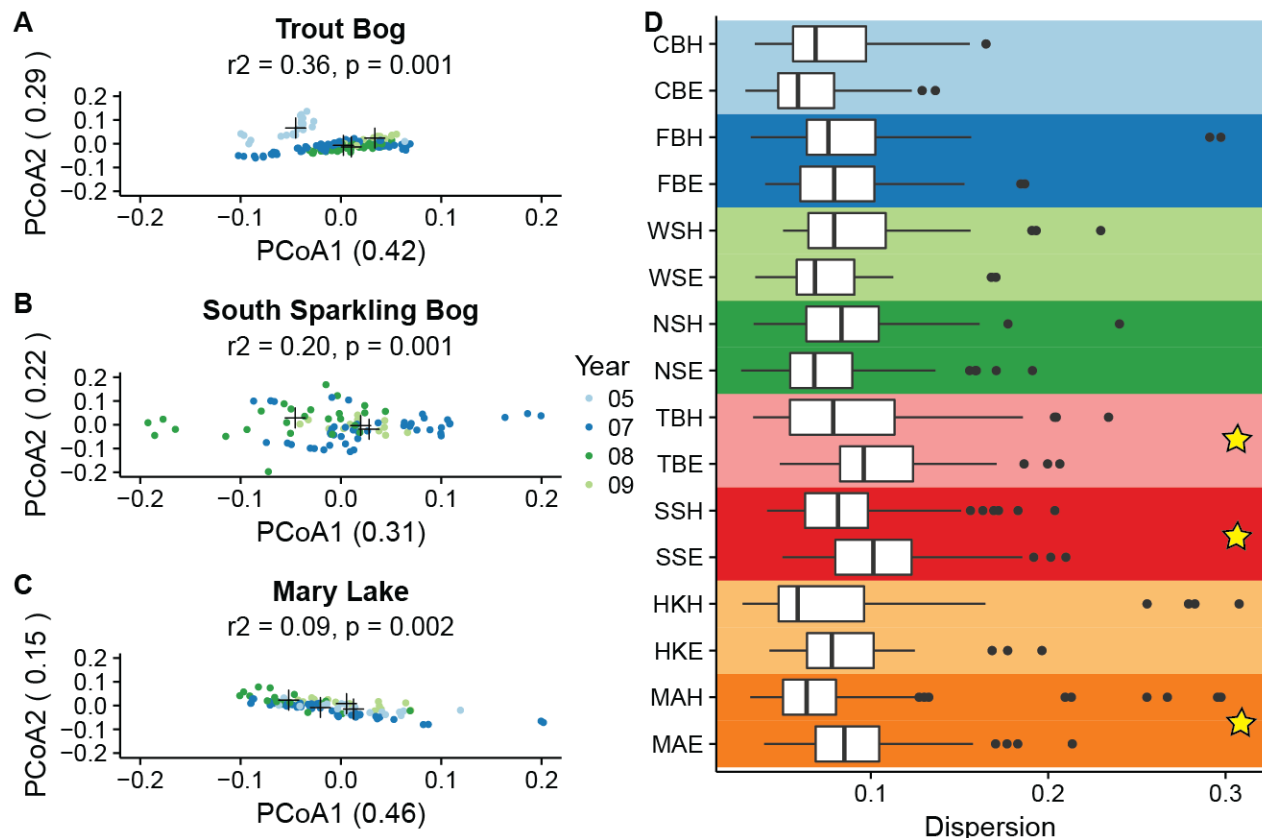
29

Figure 3. Internannual variability and dispersion by lake. Principal coordinates analysis using weighted UniFrac as the distance metric was used to measure the amount of interannual variation in the three lake hypolimnia with the longest time series (A-C). Additional ordinations of all other lakes and layers with at least two years of sampling are provided as supplemental figures (Figure S4). Black crosses indicated the centroid for each year. All hypolimnia showed significant clustering by year by PERMANOVA. Six outliers in Mary Lake from 2007 are not shown, as their coordinates lie outside the range specified for consistency between plots; these points were included in the PERMANOVA significance test. Panel D shows dispersion of each lake and layer in a PCoA including all samples (Lake abbreviations found in Table 1; E indicates epilimnion and H indicates hypolimnion; 6 outliers with distances from the centroid greater than 0.45 were removed). Stars indicate significant differences between layers at p < 0.05 by

627 PERMADISP with a Bonferroni correction for multiple pairwise comparisons. Layers were

628 signficantly different in TB, SS, and MA. No significant differences in dispersion between layers

629 in the polymictic lakes (CB, FB, and WS), meromictic lake HK, or NS, a dimictic lake with an
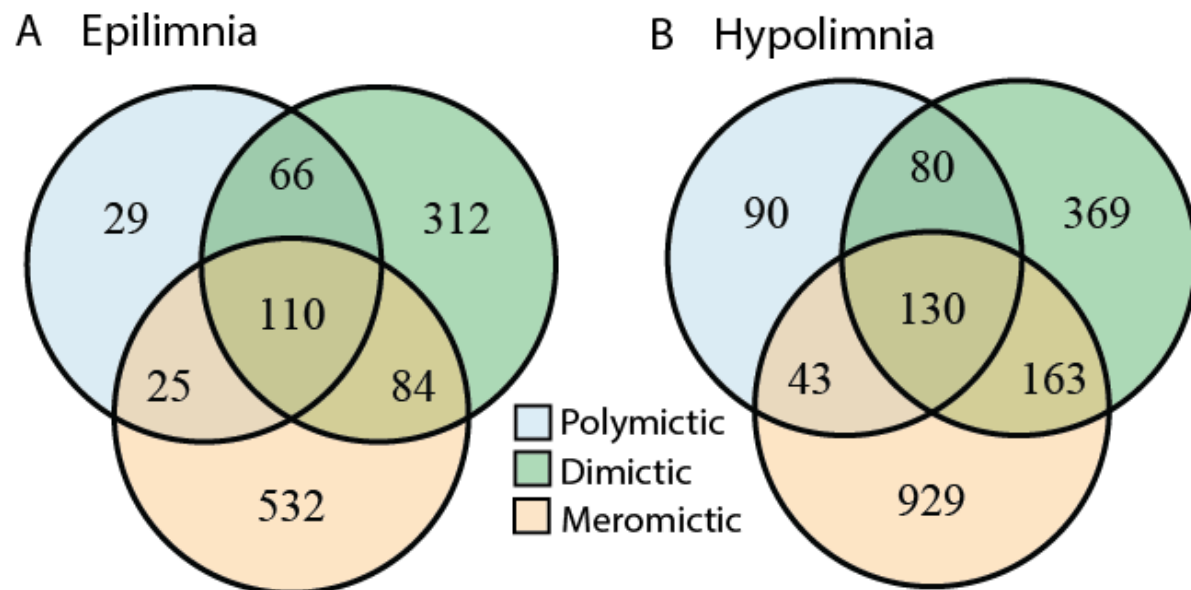
630 additional artificial mixing event.

631

632

633 **Figure 4. Numbers of unique and shared OTUs by mixing regime.** To better understand how

634 shared community membership differs by mixing regime, we quantified the number of shared

635 and unique OTUs in each category. An OTU needed only to appear in one sample at any

636 abundance to be considered present in a category. We found that in both layers, meromictic lakes

637 have the greatest numbers of unique OTUs and polymictic lakes have the least. Meromictic and

638 dimictic lakes shared the most OTUs, while meromictic and polymictic lakes shared the least.

639 Dimictic lakes shared more OTUs with meromictic lakes than with polymictic lakes.
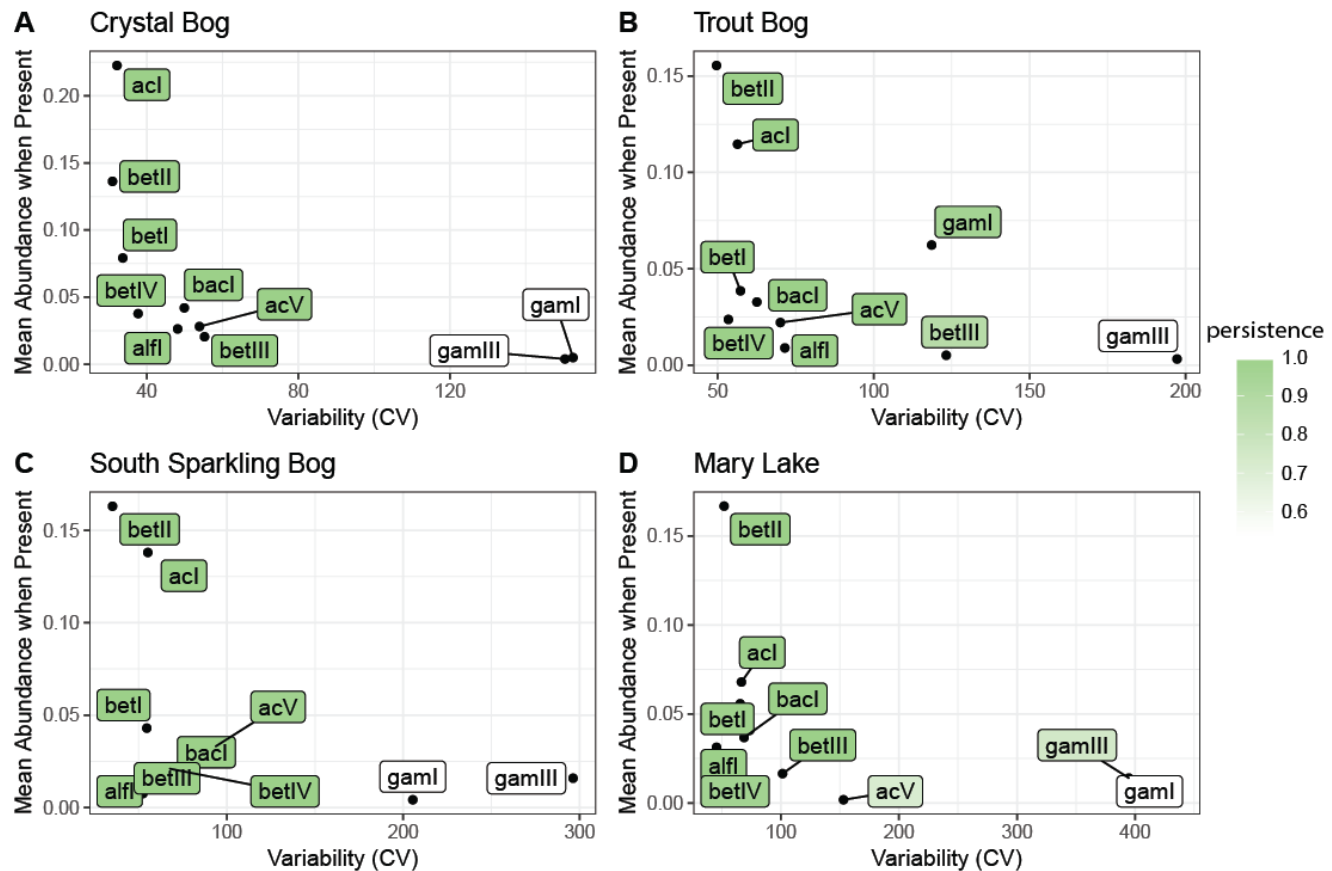
640

31

Figure 5. Traits of freshwater lineages. These well-defined freshwater groups showed similar persistence, variance, and abundance in every lake, despite differing abundance patterns. Data from epilimnia with at least two years of undisturbed sampling are shown here. Mean abundance was represented as the average percentage of reads attributed to each lineage when that lineage was present. Variability was measured as the coefficient of variation. Persistence (shaded color) was defined as the proportion of samples containing each lineage. The observed consistency in mean abundance, variability, and persistence suggests that unknown functions or metabolic characteristics drive a stable lifestyle. Additional plots by year can be found in Figure S8.