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 60637

[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: freshwater, microbial ecology, microbial communities, time series, 16S rRNA gene amplicon sequencing

Abstract word count: 232

Importance word count: 141

Article word count: 5297

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. The sites we studied encompassed a range of water column mixing frequencies, which we hypothesized would be associated with trends in alpha and beta diversity. 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 lake type sampled, including Actinobacteria in polymictic lakes, Methylophilales in dimictic lakes, and Omnitrophica in meromictic lakes. The community present in each year and site was also surprisingly unique. Despite unexpected interannual variability in community composition, we detected a core community of taxa found in all lakes and layers, including the Actinobacteria tribe acI-B2 and the Betaprotobacteria lineage PnecC. 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 multi-site observations, as analyzing only a single year of data or one lake 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 composition varies by year, a finding which likely applies to other ecosystems and has 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 global carbon cycling, and the information presented here may ultimately help refine carbon budgets for these lakes. Finally, all data and code in this study are publicly available. We hope that this will serve as a resource to anyone seeking to answer their own microbial ecology questions using a multi-year time series.

Introduction

One of the major goals of microbial ecology is to predict bacterial community composition. However, we have only a cursory knowledge of the factors that would allow us to predict bacterial community dynamics. To characterize the diversity and dynamics of an ecosystem's bacterial community, sampling the same site multiple times is just as necessary as sampling replicate ecosystems. Additionally, the sampling frequency must match the rate of change of the process being studied. We must first understand the scales on which bacterial communities change before we can design experiments that capture a full range of natural variation.

Bacterial communities have the potential to change more quickly than communities of macro-organisms due to their fast rate of reproduction. A meta-analysis of time series spanning one to three years found positive species-time relationships, indicating that more taxa are observed as the duration of sampling increases, either due to incomplete sampling, extinction and immigration, or speciation (1). Bacterial time series display time decay, meaning that the community continues to become more dissimilar from the initial sampling event as time from that event increases (2). In one freshwater lake, the amount of change in the bacterial community over a single day was equivalent to dissimilarity between sampling points ten meters apart (3). Conversely, bacterial

77 communities can also change gradually over extremely long time scales, as they are sensitive to
78 changes in environmental parameters such as nutrient availability and temperature. Wetland
79 ecosystems and their carbon emissions are expected to change on scales greater than 300 years (4);
80 as these emissions are the result of bacterial processes, we expect that the bacterial community
81 will change on the same time scale as its ecosystem. Changes in marine phytoplankton regimes
82 have been observed to occur over the past millennium, correlating with shifts in climate (5). With
83 such a large range of potential time scales for change, we now recognize the need to more
84 rigorously consider the duration and frequency of sampling in microbial ecology.

85 Multi-year studies of bacterial communities are less common due to their logistical
86 difficulties and the need for stable funding, but results from the United States National Science
87 Foundation funded Microbial Observatory and Long Term Ecological Research (LTER) projects
88 are exemplary. As a few examples among many, the San Pedro North Pacific - Microbial
89 Observatory contributed to our understanding of heterogeneity of bacterial communities across
90 space and time (6), while research at the Sapelo Island – Microbial Observatory has led the field
91 in linking genomic data to metadata (7). While there are several well-established long-term time
92 series in marine systems, studies at this scale in freshwater are rare. In our own North Temperate
93 Lakes – Microbial Observatory, based in Wisconsin, USA, a multi-year time series of
94 metagenomic data was used to study sweeps in diversity at the genome level (8), adding to our
95 knowledge of how genetic mutation influences bacterial communities. Long-term microbial
96 ecology studies have a time-tested role in the quest to forecast bacterial communities.

97 Our North Temperate Lakes - Microbial Observatory time series was collected from eight
98 bog lakes near Minocqua in the boreal region of northern Wisconsin, USA. Bog lakes contain high
99 levels of dissolved organic carbon in the form of humic and fulvic acids, resulting in dark, “tea-

colored" water. Due to their dark color, bog lakes absorb heat from sunlight, resulting in strong stratification during the summer. The top layer in a stratified lake, called the "epilimnion," is oxygen-rich and warm. At the lake bottom, a cold layer called the "hypolimnion" is formed, becoming anoxic almost immediately in darkly stained bog lakes. The transitions between mixing of these two layers and stratification occur rapidly in these systems, and at different frequencies (called mixing regimes) depending on the depth, surface area, and wind exposure of the lake. Changes in bacterial community composition along the vertical gradients established during stratification are well documented (9, 10). Mixing has been shown to be a disturbance to the bacterial communities in bog lakes (11). The bacterial communities in bog lakes are still being characterized, but contain both ubiquitous freshwater organisms (12, 13) and members of the candidate phyla radiation (14). Seasonality in freshwater lakes is thought to be the norm rather than the exception (15, 16); however, multiple years of sampling are needed to confirm these prior findings.

Our dataset is comprised of 1,387 16S rRNA gene amplicon sequencing samples, collected from eight lakes and two thermal layers over five years. Our primary goals for this dataset were to census members of the bog lake bacterial community and to identify taxa that are core to the bacterial community of bog lake ecosystems. We hypothesized that mixing regime structures the bacterial community, leading to an association between mixing frequency and alpha and beta diversity in bog lakes. Finally, we investigated seasonality at the community level, clade (i.e. roughly genus) level, and OTU level to identify annual trends. This extensive, long-term sampling effort establishes a time series that allows us to assess variability, responses to mixing frequency and re-occurring trends in freshwater bacterial communities.

Results

Overview of community composition

We used a time series of 16S rRNA gene amplicon data to investigate bacterial community composition over time and across lakes. Sampling occurred at approximately weekly intervals and primarily during the summer stratified period (May – Aug) (Figure S1). Sites were not sampled continuously over the entire time series, and metadata is available only for a subset of samples. A total of 8,795 OTUs were detected in 1,387 samples. As is typical for most freshwater ecosystems, Proteobacteria, Actinobacteria, Bacteroidetes, and Verrucomicrobia were the most abundant phyla (Figure S2). Within these phyla, OTU abundance was highly uneven. Much of the abundance of Proteobacteria could be attributed to OTUs belonging to the well-known freshwater groups *Polynucleobacter* and *Limnohabitans*, and the freshwater lineage acI contributed disproportionately to the observed abundance of Actinobacteria. Like many microbial communities, unevenness was a recurring theme in this dataset, which had a long rare tail of OTUs and trends driven largely by the most abundant OTUs (17, 18). These results show that the composition of our dataset is consistent with results from other bog lakes (10, 14).

Community richness

We hypothesized that water column mixing frequency was associated with alpha diversity. Observed richness was calculated for every sample at the OTU level, and samples were aggregated by lake and layer. Hypolimnia were generally richer than epilimnia (Figure 1, Table S1). Significant differences in richness between lakes were detected using the Wilcoxon signed rank test with a Bonferroni correction for multiple pairwise comparisons (Table S2). For both layers, polymictic lakes had the fewest taxa, dimictic lakes had intermediate numbers of taxa, and meromictic lakes had the most taxa. This dataset includes two fall mixing events (Trout Bog 2007

and North Sparkling Bog 2008), as well as the artificial mixing event in North Sparkling Bog 2008 (11). Richness decreased sharply in mixed samples compared to those taken during the summer stratified period (Figure S3). The observed association between mixing frequency and richness suggests that water column mixing (or one or more co-varying environmental parameters) structures the bacterial community.

Clusters of community composition

To determine if mixing frequency is associated with community composition, we measured beta diversity between sites, based on the relative number of reads assigned to each OTU. When differences in community composition were quantified using weighted UniFrac distance and visualized using principal coordinates analysis, several trends emerged. The large number of samples precluded much interpretation using a single PCoA, but sample clustering by layer, mixing regime, and lake was evident. Thus, we also examined PCoA for single lakes (both layers). Communities from the epilimnion and hypolimnion layers were significantly distinct from each other at $p < 0.05$ in all lakes except for polymictic Forestry Bog (FB) ($p = 0.10$) (Figure S4, a-h).

Within layers, mixing regime also explained differences in community composition (Figure 2, Table S2). Clustering by mixing regime was significant by PERMANOVA in both epilimnia and hypolimnia samples ($r2 = 0.20$ and $r2 = 0.22$, respectively, and $p = 0.001$ in both groups). Site was a strong factor explaining community composition, with significant clustering in epilimnia ($p = 0.001$, $r2 = 0.34$) and hypolimnia ($p = 0.001$, $r2 = 0.49$) (Table S2). Date and mean water temperature did not describe the observed clustering as well as lake or mixing regime (Figure S5, a-f). Because principle coordinates analysis can be susceptible to artifacts, we also performed a comparison of beta diversity between sites using a Bray-Curtis dissimilarity distance matrix without ordination; the same results were obtained (Figure S5, g-h). These findings

demonstrate that thermal layer, lake, and mixing frequency are associated with changes in bacterial community composition.

Variability and dispersion

While OTU-based community composition was distinct by layer, lake, and mixing regime, there was still variability over time. We used weighted UniFrac distance to quantify variability in beta diversity between samples within the same site and year. Each year in each lake had a significantly different community composition, indicating interannual variability in the community composition (Figure 3a-c, Figure S3, i-k, Table S2). As multiple environmental variables changed in each year of sampling, it is not clear which (if any) could explain the observed annual shifts in community composition. We found no evidence of repeating seasonal trends during the stratified summer months in these lakes in time decay plots using weighted UniFrac distance. Likewise, we examined trends in the most abundant individual OTUs and did not observed repeatable annual trends, even when abundances in each year were normalized using z-scores (Figure S6).

Varibility can also be assessed by measuring the beta diversity within a single site. We measured pairwise weighted UniFrac distance between samples in each each lake-layer (Figure 3d). This analysis showed that layers had significantly different levels of beta diversity within a single site for West Sparkling Bog, North Sparkling Bog, Trout Bog, South Sparkling Bog, and Mary Lake, as determined using a Wilcoxon signed rank test with a Bonferroni correction for multiple pairwise comparisons. Within-site beta diversity was not significantly different in Crystal Bog, Forestry Bog, and Hell's Kitchen. Mean pairwise UniFrac distance was lower in the epilimnion than the hypolimnion in the West and North Sparkling Bogs, but higher in the other three significant sites. Performing the same analysis on a single year of data with approximately

even numbers of samples from each site showed the same trends. This shows that the amount of variability in the bacterial community differs by site as well as by year.

The core community of bog lakes

One of the goals of this study was to determine the core bacterial community of bog lakes in general, and to determine if mixing regime affects core community membership. Our previous analyses showed that community composition was distinct in each layer and lake (Figure 2), while marked variability was observed within the same lake and layer (Figure 3). This prompted us to ask whether we had adequately sampled through time and space to fully census the lakes. Still, rarefaction curves generated for the entire dataset and for each layer begin to level off, suggesting that we have indeed sampled the majority of taxa found in our study sites. To identify the taxa that comprise the bog lake core community, we defined "core" as being present in 90% of a group of samples, regardless of abundance in the fully curated dataset. Core taxa are reported as OTU number and taxonomic classification our freshwater-specific database (19). Four OTUs met this criteria for all samples in the full dataset: OTU0076 (bacI-A1), OTU0097 (PnecC), OTU0813 (acI-B2), and OTU0678 (LD28). These taxa were therefore also core to both epilimnia and hypolimnia. Additional taxa core to epilimnia also included OTU0004 (betI), OTU0184 (acI-B3), OTU0472 (Lhab-A4), and OTU0522 (alfI-A1), while additional hypolimnia core taxa included OTU0042 (Rhodo), OTU0053 (unclassified Verrucomicrobia), and OTU0189 (acI-B2).

We performed the same core analysis after combining OTUs assigned to the same tribe (previously defined as sharing $\geqq$ 97% nucleotide identity in the nearly full length 16S rRNA gene and according to phylogenetic branch structure (19)) into new groups. This revealed that certain tribes were core to the entire dataset or thermal layer even though their member OTUs were specific to certain sites. Notably, some OTUs were endemic to specific lakes, even though their corresponding tribe was found in multiple lakes/layers. OTUs not classified at the tribe level were

not included. Results were similar to those observed at the OTU level, but yielded more core taxa. Tribes core to all samples included bacI-A1, PnecC, acI-B2, and LD28, but also betIII-A1 and acI-B4. In epilimnia, the core tribes were bacI-A1, PnecC, betIII-A1, acI-B3, acI-B2, Lhab-A4, alfI-A1, LD28, and acI-B4, while in hypolimnia, they were Rhodo, bacI-A1, PnecC, betIII-A1, acI-B2, and acI-B4. These results show that despite lake-to-lake differences and interannual variability, there are bacterial taxa that are consistently present in bog lakes. We note that tribes correspond very roughly to species-level designations as explained previously (19)

Principal coordinates analysis suggested that samples clustered also by mixing regime (Figure 2). We thus evaluated Venn diagrams of OTUs shared by, and unique to, each mixing regime to better visualize the overlap in community composition (Figure 4). In both epilimnia and hypolimnia, meromictic lakes had the greatest numbers of unique OTUs while polymictic lakes had the least, consistent with the differences in richness between lakes (Figure 1). Shared community membership, i.e. the number of OTUs present at any abundance in both communities, differed between mixing regimes. Epilimnia (A) and hypolimnia (B) showed similar trends in shared membership: meromictic and dimictic lakes shared the most OTUs, while meromictic and polymictic lakes shared the least.

We next used indicator analysis to identify the taxa unique to each mixing regime. Indicator analysis is a statistical method used to determine if taxa are found significantly more frequently in certain pre-determined groups of samples than in others. In this case, the groups were defined by mixing regime, and normalization was applied to account for different numbers of samples in each group. OTUs were grouped at every taxonomic level, and all taxonomic levels were run in the indictor analysis at once to account for differences in the ability of these levels to serve as indicators (for example, the order Actinomycetales is a stronger indicator of polymictic lakes than

239 the phylum Actinobacteria). An abundance threshold of 500 reads was imposed on each taxonomic
240 group. The full table of results from the indicator analysis is available as Dataset S1, while a few
241 indicator taxa of interest are highlighted here.

242 The lineage acI is a ubiquitous freshwater group, with specific clades and tribes showing a
243 preference for bog lakes in previous studies (20, 21). Our dataset shows a further distinction of acI
244 by mixing regime in epilimnia; acI-A tribes were found predominantly in meromictic lakes, with
245 exception of Phila, which is an indicator of polymictic lakes. Tribes of acI-B, particularly OTUs
246 belonging to acI-B2, were indicators of dimictic lakes. Methylophilales, a putative methylotroph,
247 was also an indicator of dimictic lakes, as was the putative sulfate reducing family
248 Desulfobulbaceae. The phyla Planctomyces, Omnitrophica (formerly OP3), OP8, and
249 Verrucomicrobia were found more often in meromictic lakes, as were putative sulfate reducing
250 taxa belonging to Syntrophobacterales and Desulfobacteraceae. Indicators of polymictic lakes
251 include ubiquitous freshwater groups such as Limnohabitans, Polynucleobacter (PnecC), betI-A,
252 and verI-A. Thus, despite the observed variability and differences between lakes, layers, and years,
253 we detected a core community composed of ubiquitous freshwater bacterial groups. Additionally,
254 we identified indicator taxa endemic to groups of sites defined by mixing frequency.

255 **Lifestyles of freshwater lineages**

256 Because of the observed variability in bacterial community dynamics, we next asked if
257 individual OTUs showed consistent levels of abundance, persistence, and variability. We defined
258 these metrics as mean abundance when present, the proportion of samples containing the group of
259 interest, and the coefficient of variation for lineages classified using the freshwater taxonomy,
260 respectively. These metrics have been previously used to categorize OTUs (22, 23). Using only
261 well-defined freshwater groups allowed better taxonomic resolution as we summed the abundances
262 of OTUs by their lineage classification. We note that lineage is very roughly analogous to family

in our provisional freshwater taxonomy (19). Lifestyle traits of lineages were consistent across both lakes and years. Low persistence was associated with high variability, and low variability was associated with high abundance (Figure 5, Figure S7). We rarely observed "bloomers," situations where a clade had both high abundance and low persistence; one potential reason for this could be that true "bloomers" drop below the detection limit of our sequencing methods when not abundant. Most freshwater lineages were highly persistent at low abundances with low variability. Lineage gamIII of the Gammaproteobacteria was an exception, with low persistence, low abundance, and high variability. Lineages gamI and verI-A occasionally also exhibited this profile. Lineages betII and acI were highly abundant and persistent with low variability, consistent with their suggested lifestyles as ubiquitous freshwater generalists (12, 21). Even though OTUs did not show the same abundance dynamics each year, they did exhibit patterns that are consistent between years and lakes.

Discussion

The North Temperate Lakes - Microbial Observatory bog dataset is a comprehensive 16S rRNA gene amplicon survey spanning four years, eight lakes, and two thermal layers. We hypothesized that alpha and beta diversity would be associated with mixing frequency in bog lakes. Richness and membership in these communities were structured by layer, mixing regime, and lake. However, we found that multiple years of sampling were necessary to census the community of bog lake ecosystems. We identified specific bacterial taxa core to bog lakes, as well as taxa endemic to certain depths or mixing regimes. High levels of variability were detected in this dataset; the community composition observed in each lake and each year of sampling was unique. However, freshwater lineages still showed consistent lifestyles, defined by abundance, persistence, and variability, across lakes and years, even though the abundance trends of individual OTUs

varied each year. Our results emphasize the importance of multiple sampling events to assess full bacterial community membership and variability in an ecosystem.

The bog lakes in this study have been model systems for freshwater microbial ecology for many years. Early studies used Automated Ribosomal Intergenic Spacer Analysis (ARISA), a fingerprinting technique for identifying unique bacterial taxa in environmental samples (24). Our research built upon these studies and added information about the taxonomic identities of bacterial groups. For example, persistent and unique bacterial groups were detected in the bog lakes using ARISA (25); using 16S rRNA gene amplicon sequencing, we also found persistent groups and could identify them as the ubiquitous freshwater bacteria LD28, acI-B2, PnecC, and bacI-A1. Differences in richness and community membership were previously detected within one year, between Crystal Bog, Trout Bog, and Mary Lake, three sites representative of the three mixing regime categories of polymictic, dimictic, and meromictic (25). Our data supported these results and suggest that these trends are indeed linked with mixing regime, as we included multiple lakes of each type sampled over multiple years in this study.

We supported previous research on the characteristics of bacterial communities in the epilimnion and hypolimnion and the association of lake mixing frequency with community composition. We confirmed that epilimnia communities tended to be more dispersed than hypolimnia communities, potentially due to increased exposure to climatic events (25). Mixing was disruptive to both epilimnion and hypolimnion communities, selecting for only a few taxa that persist during this disturbance, but quickly recovering diversity once stratification was re-established (11, 26). Our initial inspiration for the collection of this dataset was the intermediate disturbance hypothesis. We hypothesized that water column mixing is a disturbance to bog lake bacterial communities, and that lakes with intermediate mixing frequency would have the highest

levels of biodiversity. Comparing richness between lakes of different mixing regimes did not support the intermediate disturbance hypothesis; rather, the least frequently mixing lakes had the most diverse communities. Richness also correlated positively with lake volume, a potential result of a positive taxa-area relationship, but more lakes of similar volumes with varying depths are needed to prove this relationship in our study system (27, 28). As many variables co-vary with volume, including mixing frequency and concentrations of nitrogen and dissolved carbon, we cannot determine which of these factors lead to the observed differences in diversity between sites based on our dataset.

We were not able to detect repeatable annual trends in bog lakes in our multiple years of sampling. While seasonality in marine and river systems has been well-established by our colleagues, previous research on seasonality in freshwater lakes has produced inconsistent results (29–32). Distinct, seasonally repeatable community types were identified in alpine lakes, but stratified summer communities were distinct each year (33). Seasonal trends were detected in a time series from Lake Mendota similar to this study, but summer samples in Lake Mendota were more variable then those collected in other seasons (34). In the previous ARISA-based research on the bog lakes in our dataset, community properties such as richness and rate of change were consistent each year, and the phytoplankton communities were hypothesized to drive seasonal trends in the bacterial communities based on correlation studies (35–37). Synchrony in seasonal trends was observed (36); however, in a second year of sampling for seasonal trends in Crystal Bog and Trout Bog, these findings were not reproduced (38). Successional trends were studied in Crystal Bog and Lake Mendota with a relatively small number of samples collected over two years and “dramatic changes” in community composition associated with drops in biodiversity were described during the summer months, while spring, winter, and fall had more stable community composition (35).

Because our dataset was sparsely represented by seasons other than summer, higher summer variability may explain why we see a different community each year and a lack of seasonal trends in community composition. However, we cannot disprove the influence of seasonality on bacterial community dynamics in temperate freshwater lakes as a general feature. Our results may indeed point to a feature that is unique to darkly stained acidic bog lakes. Even in marine systems, trends in seasonality differ by site and OTU definition, and continued long term time series sampling is suggested as an approach needed to elucidate these trends and link seasonality in bacterial community composition to biogeochemical cycling (39).

One of the biggest benefits of 16S rRNA gene amplicon sequencing over ARISA is the ability to assign taxonomic classifications to sequences. Tracking bacterial taxa through multiple sites and over multiple years allowed us to detect consistent lifestyle trends, despite a lack of predictability in seasonal trends. Some groups, such as acI (Actinobacteria) and betII (Betaproteobacteria), were persistent, abundant, and not variable, much like *Pelagibacter ubique* (SAR11) in marine systems. Other freshwater taxa such as gamI and gamIII (both Gammaproteobacteria) exhibited a pattern of low persistence, low abundance, and high variability. Unlike in the oceans, where taxa such as *Alteromonas* "bloom and bust" (40), no taxa classified within the freshwater taxonomy with high abundance and low persistence or high variability were observed. This suggest that either bog lakes are not conducive to the large blooms of a single population as observed in other freshwater lakes, or that taxa with this lifestyle dropped below our detection limit when not blooming.

In addition to a core of persistent taxa found in nearly every sample collected, we also identified taxa endemic to either the epilimnion or hypolimnion and to specific mixing regimes. These endemic taxa likely reflect the physical and/or biogeochemical differences driven by mixing

355 regime. Dimictic and meromictic hypolimnia, which are consistently anaerobic, harbor putative
356 sulfur cycling groups not present in polymictic hypolimnia, which are more frequently oxygenated.
357 Members of the acI lineage partition by mixing regime in epilimnia, and the functional traits
358 driving this filtering effect are the subject of active study (20). Interestingly, the meromictic Mary
359 Lake hypolimnion contains several taxa classified into the candidate phyla radiation (41) and a
360 larger proportion of completely unclassified reads than other hypolimnia. This is consistent with
361 the findings of other 16S rRNA gene amplicon sequencing and metagenomics studies of
362 meromictic lakes (42, 43), and suggests that the highly reduced and consistently anaerobic
363 conditions in meromictic hypolimnia are excellent study systems for research on members of the
364 candidate phyla radiation and "microbial dark matter".

365 Perhaps the biggest implication of this study is the importance of repeated sampling of
366 microbial ecosystems. A similar dataset spanning only a single year would not have captured the
367 full extent of variability observed, and therefore would not have detected as many of the taxa
368 belonging to the bog lake community; even our four years of weekly sampling during the summer
369 stratified period did not result in level rarefaction curves (Figure S7). While we found no evidence
370 for seasonal trends or repeated annual trends, it is possible that there are cycles or variables acting
371 on scales longer than the five years covered in this dataset, or that interannual differences are
372 driven by environmental factors that do not occur every year. Unmeasured biotic interactions
373 between bacterial taxa may also contribute to the observed variability. Understanding the factors
374 that contribute to variability in lake communities will lead to improved predictive modelling in
375 freshwater systems, allowing forecasting of bloom events and guiding better management
376 strategies. Additionally, these systems may be ideal for addressing some of the core questions in

377 microbial ecology, such as how community assembly occurs, how interactions between taxa shape
378 community composition, and how resource partitioning drives the lifestyles of bacterial taxa.
379 To answer these questions and more, we continue to collect and sequence samples for the North
380 Temperate Lakes – Microbial Observatory, and we are expanding our sequencing repertoire
381 beyond 16S rRNA gene amplicon sequencing. All data we have currently generated can be found
382 in the R package "OTUtable" which is available on CRAN for installation via the R command
383 line, or on our GitHub page. This dataset has already been used in a meta-analysis of microbial
384 time series (1). We hope that this dataset and its future expansion will be used as a resource for
385 researchers investigating their own questions about how bacterial communities behave on long
386 time scales.
387

388 # Materials and Methods

389 **Sample Collection**

390 Water was collected from eight bog lakes during the summers of 2005, 2007, 2008 and
391 2009, as previously described (25). Briefly, the epilimnion and hypolimnion layers were collected
392 separately using an integrated water column sampler. Dissolved oxygen and temperature profiles
393 were measured at the time of collection using a handheld YSI 550A (YSI Inc., Yellow Springs,
394 OH). After transport to the laboratory, two biological replicates were taken by filtering
395 approximately 150 mL from each well-mixed sample through 0.22 micron polyethersulfone filters
396 (Supor 200, Pall, Port Washington, NY). Filters were stored at -80C until DNA extraction using
397 FastDNA Spin Kit for Soil (MP Biomedicals, Santa Ana, CA), with minor modifications (44). The
398 sampling sites are located near Boulder Junction, WI, and were chosen to include lakes represent
399 the three mixing regimes of polymictic (multiple mixing events per year), dimictic (two mixing

events per year, usually in spring and fall), and meromictic (no recorded mixing events) (Table 1). Trout Bog and Crystal Bog are also primary study sites for the North Temperate Lakes - Long Term Ecological Research Program (NTL-LTER), which measures a suite of chemical limnology parameters fortnightly during the open water season. The NTL-LTER also maintains autonomous sensing buoys on Trout Bog and Crystal Bog, allowing for more refined mixing event detection based on thermistor chain measurements.

Sequencing

A total of 1,510 DNA samples, including 547 biological replicates, were sequenced by the Earth Microbiome Project according to their standard protocols in 2010, using the original V4 primers (FWD:GTGCCAGCMGCCGCGGTAA; REV:GGACTACHVGGGTWTCTAAT) (45). Briefly, the V4 region was amplified and sequenced using Illumina HiSeq, resulting in 77,517,398 total sequences with an average length of 150 base pairs. To reduce the number of erroneous sequences, QIIME's "deblurring" algorithm for reducing sequence error in Illumina data was applied (46). Based on the sequencing error profile, this algorithm removes reads that are likely to be sequencing errors if those reads are both low in abundance and highly similar to a high abundance read. Reads occurring less than 25 times in the entire dataset were removed after deblurring, leaving 9,856 unique sequences. These sequences are considered operational taxonomic units (OTUs).

570 sequences with long homopolymer runs, ambiguous base calls, or incorrect sequence lengths were found and removed via mothur v1.34.3 (47). Thirty-three chimeras and 340 chloroplast sequences (based on pre-clustering and classification with the Greengenes 16S rRNA gene database, May 2013) (48) were removed. Samples were rarefied to 2,500 reads; samples with less than 2,500 reads were omitted, resulting in 1,387 remaining samples. The rarefaction cutoff used was determined based on the results of simulation; 2,500 reads was chosen to maximize the

number of samples retained, while maintaining sufficient quality for downstream analysis of diversity metrics.

Representative sequences for each OTU were classified in either our curated freshwater database (19) or the Greengenes database based on the output of NCBI-BLAST (blast+ 2.2.3.1) (49). Representative sequences from each OTU were randomly chosen. The program blastn was used to compare representative sequences to full-length sequences in the freshwater database. OTUs matching the freshwater database with a percent identity greater than 98% were classified in that database, and remaining sequences were classified in the Greengenes database. Both classification steps were performed in mothur using the Wang method (50), and classifications with less than 70% confidence were not included. A detailed workflow for quality control and classification of our sequences is available at (https://github.com/McMahonLab/16STaxAss) (manuscript in prep).

Statistics

Statistical analysis was performed in R v3.3.2 (R Development Core Team, 2008. R: A language and environment for statistical computing.). Significant differences in richness between lakes was tested using a pairwise Wilcoxon sum rank test with a Bonferroni adjustment in the R package "exactRankTests" (T. Hothorn and K. Hornik, 2015. exactRankTests: Exact Distributions for Rank and Permutation Tests). Similarity between samples was compared using weighted UniFrac distance, implemented in "phyloseq" (51) (P.J. McMurdie and S. Holmes, 2013. phyloseq: An R Package for reproducible interactive analysis and graphic of microbiome census data). Weighted UniFrac distance was chosen because it explained the greatest amount of variation in the first two axes of a principle coordinates analysis, performed in "vegan" (J. Oksanen, 2016. vegan: Community Ecology Package). Other metrics tested included unweighted UniFrac distance, Bray-Curtis Dissimilarity, and Jaccard Similarity; the output of all metrics were

448 correlated. Significant clustering by year in PCoA and in dispersion between lakes was tested
449 using PERMANOVA with the function adonis() in "vegan." Trimming of rare taxa did not impact
450 the clustering observed in ordinations, such as those present in Figure 2, even when taxa observed
451 less than 1000 times were removed.

452 Indicator species analysis was performed using "indicspecies" (52). Only taxa with read
453 abundances of at least 500 reads in the entire dataset were used for this analysis. The group-
454 normalized coefficient of correlation was chosen for this analysis because it measures both positive
455 and negative habitat preferences and accounts for differences in the number of samples from each
456 site. All taxonomic levels were included in this analysis to determine which level of resolution was
457 the best indicator for each taxonomic group.

458 Plots were generated using "ggplot2" (H. Wickham, 2009. ggplot2: Elegant Graphics for
459 Data Analysis) and "cowplot" (C. Wilke, 2016. cowplot: Streamlined Plot Themes and Plot
460 Annotations for 'ggplot2'). "reshape2" was used for data formatting (H. Wickham, 2007.
461 Reshaping Data with the reshape Package).

462 **Data Availability**

463 Data and code from this study can be downloaded from the R package "OTUtable"
464 available through the Comprehensive R Archive Network (cran.r-project.org), which can be
465 accessed via the R command line using install.packages("OTUtable"), and from the McMahon
466 Lab GitHub repository "North_Temperate_Lakes-Microbial_Observatory"
467 (github.com/McMahonLab/North_Temperate_Lakes-Microbial_Observatory). Raw sequence
468 data is available through QIITA (http://qiita.microbio.me) and the European Bioinformatics
469 Institute (http://www.ebi.ac.uk/) at accession number ERP016854.

470

Acknowledgments

This study was made possible through generous support by the Earth Microbiome Project, which is funded in part through awards by the W.M. Keck and Templeton Foundations. AL is supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. (DGE-1256259). KDM acknowledges funding from the United States National Science Foundation Microbial Observatories program (MCB-0702395), the Long Term Ecological Research program (NTL-LTER DEB-1440297) and an INSPIRE award (DEB-1344254). This work was supported in part by the U.S. Dept. of Energy under Contract DE-AC02-06CH11357.

We thank the North Temperate Lakes Microbial Observatory 2005, 2007, 2008, and 2009 field crews, UW-Trout Lake Station, the UW Center for Limnology, and the Global Lakes Ecological Observatory Network for field and logistical support. Special thanks to Sara Paver and Sara Yeo for coordinating field crews in 2009. We thank Greg Caporaso for contributions during early stages of data analysis and Amnon Amir for early access to the deblurring algorithm. We thank McMahon Lab members Robin Rohwer and Joshua Hamilton for early access to a workflow used to assign taxonomies to OTUs using a custom 16S rRNA gene amplicon training set. We acknowledge efforts by many McMahon Lab undergrads and technicians related to sample collection and DNA extraction, particularly Georgia Wolfe. Finally, we personally thank the individual program directors and leadership at the National Science Foundation for their commitment to continued support of long term ecological research.

References

1. **Shade A**, **Caporaso JG**, **Handelsman J**, **Knight R**, **Fierer N**. 2013. A meta-analysis of changes in bacterial and archaeal communities with time. ISME J **7**:1493–506.

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

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

3. **Jones SE**, **Cadkin TA**, **Newton RJ**, **McMahon KD**. 2012. Spatial and temporal scales of aquatic bacterial beta diversity. Front Microbiol **3**:1–10.

4. **Mitsch WJ**, **Bernal B**, **Nahlik AM**, **Mander Ü**, **Zhang L**, **Anderson CJ**, **Jørgensen SE**, **Brix H**. 2013. Wetlands, carbon, and climate change. Landsc Ecol **28**:583–597.

5. **McMahon KW**, **McCarthy MD**, **Sherwood OA**, **Larsen T**, **Guilderson TP**. 2015. Millennial-scale plankton regime shifts in the subtropical North Pacific Ocean. Science (80-) **350**:1530–1533.

6. **Hewson I**, **Steele JA**, **Capone DG**, **Fuhrman JA**. 2006. Remarkable heterogeneity in meso- and bathypelagic bacterioplankton assemblage composition. Limnol Oceanogr **51**:1274–1283.

7. **Gifford SM**, **Sharma S**, **Moran MA**. 2014. Linking activity and function to ecosystem dynamics in a coastal bacterioplankton community. Front Microbiol **5**:1–12.

8. **Bendall ML**, **Stevens SLR**, **Chan L**, **Malfatti S**, **Schwientek P**, **Tremblay J**, **Schackwitz W**, **Martin J**, **Pati A**, **Bushnell B**, **Froula J**, **Kang D**, **Tringe SG**, **Bertilsson S**, **Moran MA**, **Shade A**, **Newton RJ**, **McMahon KD**, **Malmstrom RR**. 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J **10**:1589–1601.

9. **Taipale S**, **Jones R**, **Tiirola M**. 2009. Vertical diversity of bacteria in an oxygen-stratified humic lake, evaluated using DNA and phospholipid analyses. Aquat Microb Ecol **55**:1–16.

10. **Garcia SL**, **Salka I**, **Grossart HP**, **Warnecke F**. 2013. Depth-discrete profiles of bacterial communities reveal pronounced spatio-temporal dynamics related to lake

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

11. **Shade A**, **Read JS**, **Youngblut ND**, **Fierer N**, **Knight R**, **Kratz TK**, **Lottig NR**, **Roden EE**, **Stanley EH**, **Stombaugh J**, **Whitaker RJ**, **Wu CH**, **McMahon KD**. 2012. Lake microbial communities are resilient after a whole-ecosystem disturbance. ISME J **6**:2153–2167.

12. **Hahn MW**, **Scheuerl T**, **Jezberová J**, **Koll U**, **Jezbera J**, **Šimek K**, **Vannini C**, **Petroni G**, **Wu QL**. 2012. The passive yet successful way of planktonic life: Genomic and experimental analysis of the ecology of a free-living polynucleobacter population. PLoS One **7**:1–17.

13. **Garcia SL**, **McMahon KD**, **Grossart HP**, **Warnecke F**. 2013. Successful enrichment of the ubiquitous freshwater acI Actinobacteria. Environ Microbiol Rep 1–7.

14. **Peura S**, **Eiler A**, **Bertilsson S**, **Nyka H**, **Tiirola M**, **Jones RI**. 2012. Distinct and diverse anaerobic bacterial communities in boreal lakes dominated by candidate division OD1. ISME J **6**:1640–1652.

15. **Eiler A**, **Heinrich F**, **Bertilsson S**. 2012. Coherent dynamics and association networks among lake bacterioplankton taxa. ISME J **6**:330–42.

16. **Graham JM**, **Kent AD**, **Lauster GH**, **Yannarell AC**, **Graham LE**, **Triplett EW**. 2004. Seasonal dynamics of phytoplankton and planktonic protozoan communities in a northern temperate humic lake: Diversity in a dinoflagellate dominated system. Microb Ecol **48**:528–540.

17. **Mariadassou M**, **Pichon S**, **Ebert D**. 2015. Microbial ecosystems are dominated by specialist taxa. Ecol Lett **18**:974–982.

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

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

19. **Newton RJ**, **Jones SE**, **Eiler A**, **McMahon KD**, **Bertilsson S**. 2011. A guide to the natural history of freshwater lake bacteria. Microbiol Mol Biol Rev **75**:14–49.

20. **Garcia SL**, **Buck M**, **McMahon KD**, **Grossart H-P**, **Eiler A**, **Warnecke F**. 2015. Auxotrophy and intra-population complementary in the "interactome" of a cultivated freshwater model community. Mol Ecol **24**:4449–4459.

21. **Ghylin TW**, **Garcia SL**, **Moya F**, **Oyserman BO**, **Schwientek KT**, **Mutschler J**, **Dwulit-Smith J**, **Chan L-K**, **Martinez-Garcia M**, **Sczyrba A**, **Stepanauskas R**, **Grossart H-P**, **Woyke T**, **Warnecke F**, **Malmstrom R**, **Bertilsson S**, **McMahon KD**. 2014. Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. ISME J **8**:2503–2516.

22. **Herren CM**, **Webert KC**, **McMahon KD**. 2016. Environmental Disturbances Decrease the Variability of Microbial Populations within Periphyton. mSystems **1**:1–14.

23. **Shade AL**, **Gilbert JA**. 2015. Temporal patterns of rarity provide a more complete view of microbial diversity. Trends Microbiol **23**:335–340.

24. **Fisher MM**, **Triplett EW**. 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. Appl Environ Microbiol **65**:4630–4636.

25. **Shade A**, **Jones SE**, **McMahon KD**. 2008. The influence of habitat heterogeneity on freshwater bacterial community composition and dynamics. Environ Microbiol **10**:1057–1067.

26. **Shade A**, **Read JS**, **Welkie DG**, **Kratz TK**, **Wu CH**, **McMahon KD**. 2011. Resistance, resilience and recovery: Aquatic bacterial dynamics after water column disturbance.

Environ Microbiol **13**:2752–2767.

27. **Prosser JI**, **Bohannan BJM**, **Curtis TP**, **Ellis RJ**, **Firestone MK**, **Freckleton RP**, **Green JL**, **Green LE**, **Killham K**, **Lennon JJ**, **Osborn AM**, **Solan M**, **Gast CJ Van Der**, **Young JPW**. 2007. The role of ecological theory in microbial ecology. Nat Rev Microbiol **5**:384–392.

28. **Martiny JBH**, **Bohannan BJM**, **Brown JH**, **Kane M**, **Krumins JA**, **Kuske CR**, **Morin PJ**, **Naeem S**, **Øvreås L**, **Reysenbach A-L**, **Smith VH**. 2006. Microbial biogeography : putting microorganisms on the map. Nat Rev Genet **4**:102–112.

29. **Crump BC**, **Hobbie JE**. 2005. Synchrony and seasonality in bacterioplankton communities of two temperate rivers. Limnol Oceanogr **50**:1718–1729.

30. **Gilbert JA**, **Steele JA**, **Caporaso JG**, **Steinbruck L**, **Reeder J**, **Temperton B**, **Huse S**, **McHardy AC**, **Knight R**, **Joint I**, **Somerfield P**, **Fuhrman JA**, **Field D**. 2012. Defining seasonal marine microbial community dynamics. ISME J **6**:298–308.

31. **Fuhrman JA**, **Hewson I**, **Schwalbach MS**, **Steele JA**, **Brown M V**, **Naeem S**. 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proc Natl Acad Sci USA **103**:13104–13109.

32. **Cram J**, **Chow C**, **Sachdeva R**, **Needham D**, **Parada A**, **Steele J**, **Fuhrman J**. 2015. Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. ISME J **9**:563–580.

33. **Nelson CE**. 2009. Phenology of high-elevation pelagic bacteria: the roles of meteorologic variability, catchment inputs and thermal stratification in structuring communities. ISME J **3**:13–30.

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

dynamics and co-occurrences within freshwater bacterioplankton communities from eutrophic Lake Mendota, WI, USA. ISME J **7**:680–4.

35. **Yannarell AC**, **Kent AD**, **H. LG**, **Kratz TK**, **Triplett EW**. 2003. Temporal Patterns in Bacterial Communities in Three Temperate Lakes of Different Trophic Status. Microb Ecol **46**:391–405.

36. **Kent AD**, **Yannarell AC**, **Rusak JA**, **Triplett EW**, **McMahon KD**. 2007. Synchrony in aquatic microbial community dynamics. ISME J **1**:38–47.

37. **Kent AD**, **Jones SE**, **Lauster GH**, **Graham JM**, **Newton RJ**, **McMahon KD**. 2006. Experimental manipulations of microbial food web interactions in a humic lake: Shifting biological drivers of bacterial community structure. Environ Microbiol **8**:1448–1459.

38. **Rusak JA**, **Jones SE**, **Kent AD**, **Shade A**, **McMahon TD**. 2009. Spatial synchrony in microbial community dynamics : testing among-year and lake patterns. Verh Internat Verein Limnol **30**:936–940.

39. **Giovannoni SJ**, **Vergin KL**. 2012. Seasonality in Ocean Microbial Communities. Science (80-) **335**:671–677.

40. **Ivars-Martinez E**, **Auria GD**, **Mira A**, **Ferriera S**, **Johnson J**, **Friedman R**, **Rodriguez-Valera F**. 2008. Comparative genomics of two ecotypes of the marine planktonic copiotroph Alteromonas macleodii suggests alternative lifestyles associated with different kinds of particulate organic matter. ISME J **2**:1194–1212.

41. **Hug LA**, **Baker BJ**, **Anantharaman K**, **Brown CT**, **Probst AJ**, **Castelle CJ**, **Butterfield CN**, **Hernsdorf AW**, **Amano Y**, **Ise K**, **Suzuki Y**, **Dudek N**, **Relman DA**, **Finstad KM**, **Amundson R**, **Thomas BC**, **Banfield JF**. 2016. A new view of the tree of life. Nat Microbiol **1**:1–6.

42. **Gies EA**, **Konwar KM**, **Beatty JT**, **Hallam SJ**. 2014. Illuminating microbial dark matter in meromictic Sakinaw Lake. Appl Environ Microbiol **80**:6807–6818.

43. **Borrel G**, **Lehours AC**, **Bardot C**, **Bailly X**, **Fonty G**. 2010. Members of candidate divisions OP11, OD1 and SR1 are widespread along the water column of the meromictic Lake Pavin (France). Arch Microbiol **192**:559–567.

44. **Shade A**, **Kent AD**, **Jones SE**, **Newton RJ**, **Triplett EW**, **McMahon KD**. 2007. Interannual dynamics and phenology of bacterial communities in a eutrophic lake. Limnol Oceanogr **52**:487–494.

45. **Caporaso JG**, **Lauber CL**, **Walters WA**, **Berg-Lyons D**, **Huntley J**, **Fierer N**, **Owens SM**, **Betley J**, **Fraser L**, **Bauer M**, **Gormley N**, **Gilbert JA**, **Smith G**, **Knight R**. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J **6**:1621–1624.

46. **Amir A**, **McDonald D**, **Navas-Molina J**, **Kopylova E**, **Morton J**, **Xu Z**, **Kightley E**, **Thompson L**, **Hyde E**, **Gonzalez A**, **Knight R**. 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems **2**:1–7.

47. **Schloss PD**, **Westcott SL**, **Ryabin T**, **Hall JR**, **Hartmann M**, **Hollister EB**, **Lesniewski RA**, **Oakley BB**, **Parks DH**, **Robinson CJ**, **Sahl JW**, **Stres B**, **Thallinger GG**, **Van Horn DJ**, **Weber CF**. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol **75**:7537–7541.

48. **DeSantis TZ**, **Hugenholtz P**, **Larsen N**, **Rojas M**, **Brodie EL**, **Keller K**, **Huber T**, **Dalevi D**, **Hu P**, **Andersen GL**. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol **72**:5069–5072.

49. **Camacho C**, **Coulouris G**, **Avagyan V**, **Ma N**, **Papadopoulos J**, **Bealer K**, **Madden TL**. 2009. BLAST plus : architecture and applications. BMC Bioinformatics **10**:1–9.

50. **Wang Q**, **Garrity GM**, **Tiedje JM**, **Cole JR**. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol **73**:5261–5267.

51. **Lozupone C**, **Knight R**. 2005. UniFrac : a new phylogenetic method for comparing microbial communities UniFrac : a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol **71**:8228–8235.

52. **De Cáceres M**, **Legendre P**. 2009. Associations between species and groups of sites: indices and statistical inference. Ecology **90**:3566–3574.

Tables

Table 1. Location and characteristics of study sites. The lakes included in this time series are small, humic bog lakes in the boreal region near Minocqua, Wisconsin, USA. They range in depth from 2 to 21.5 meters and encompass a range of water column mixing frequencies (termed regimes). Dimictic lakes mix twice per year, typically in fall and spring, while polymictic lakes can mix more than twice throughout the spring, summer, and fall. Meromictic lakes have no recorded mixing events. pH was measured in 2007, while nutrient data was measured in 2008 (with the exceptions of FB, WS, and HK, measured in 2007); both measurements were taken concurrently with the bacterial biomass collection from the same water sample. Standard deviations for DOC/DIC are reported in parentheses. When two values are present in a single box, the first represents the epilimnion value and the second represents the hypolimnion value.

	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	2.5	4.5	4.6	7	8	19.3	21.5
Surface area *(m^2)*	1300	5600	4700	11900	10100	4400	30000	12000
Approx. Volume (m^3)	867	4667	7050	18247	23567	11733	193000	86000
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
Std Dev	(0.28, 1.17)	(0.15, 0.50)	(0.23, 0.72)	(0.17, 0.36)	(0.66, 54)	(0.24, 1.56)	(0.35, 1.03)	(5.66, 7.69)
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
Std Dev	(0.59, 0.10)	(4.12, 0.82)	(1.16, 0.96)	(0.43, 0.73)	(2.76, 1.17)	(0.38, 4.76)	(1.03, 0.12)	(1.91, 72.67)
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

664

665

Figures

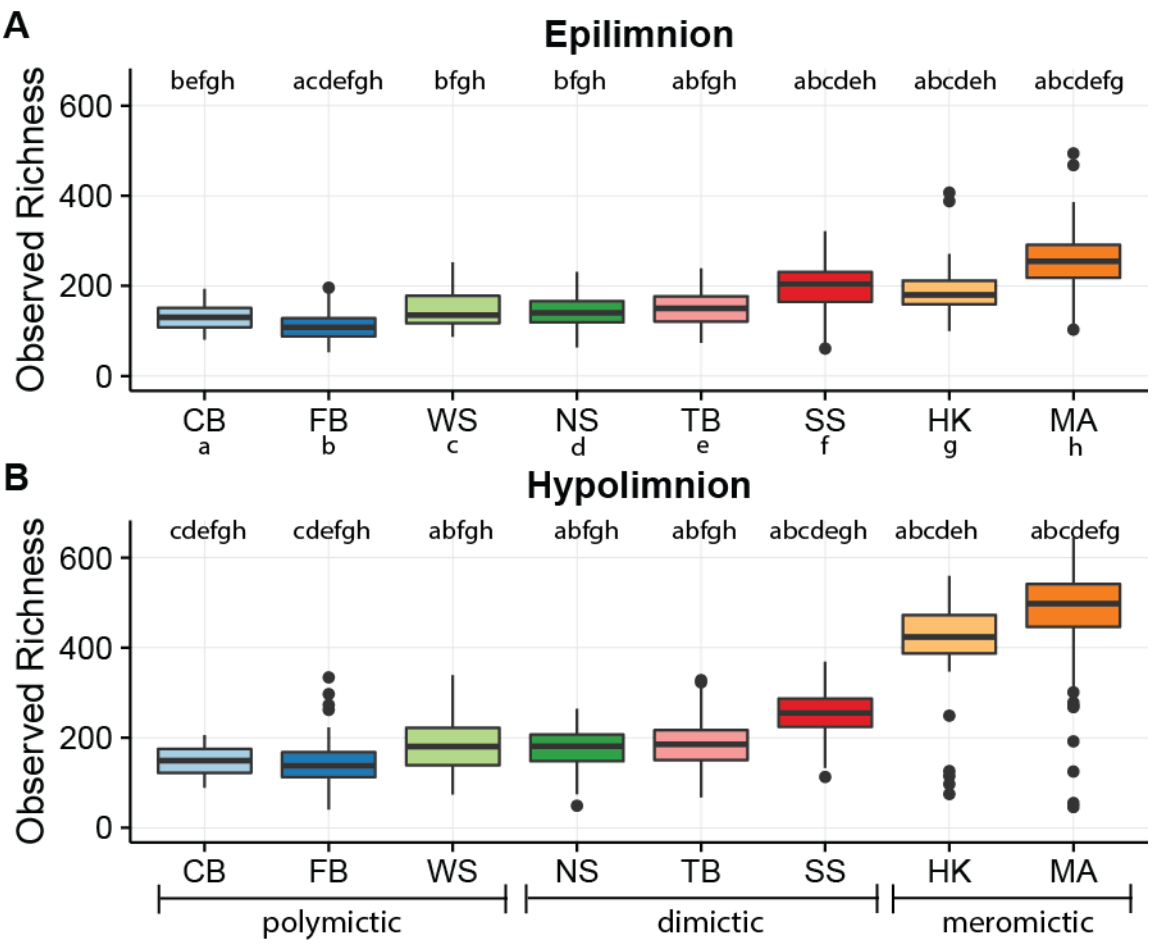
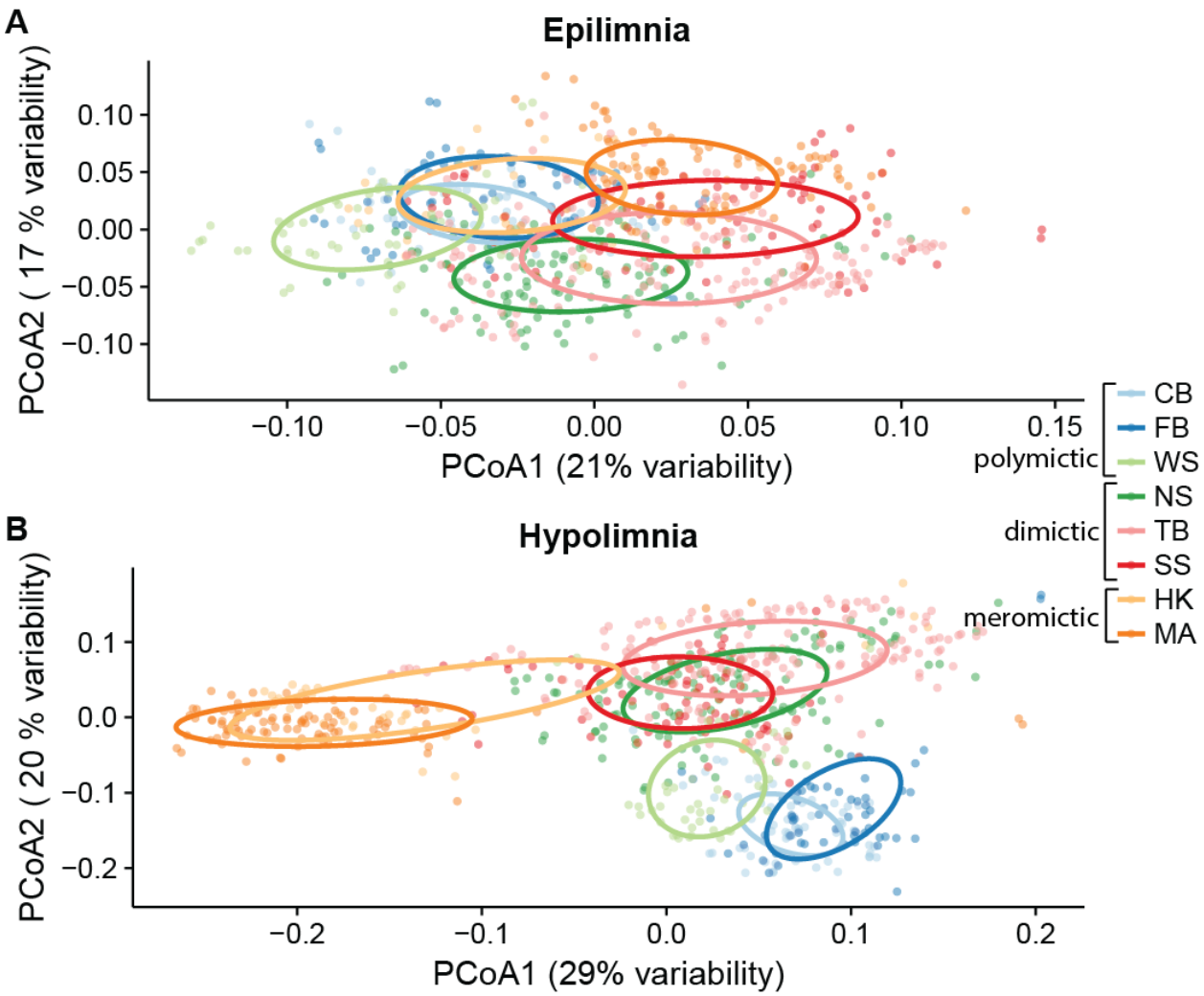


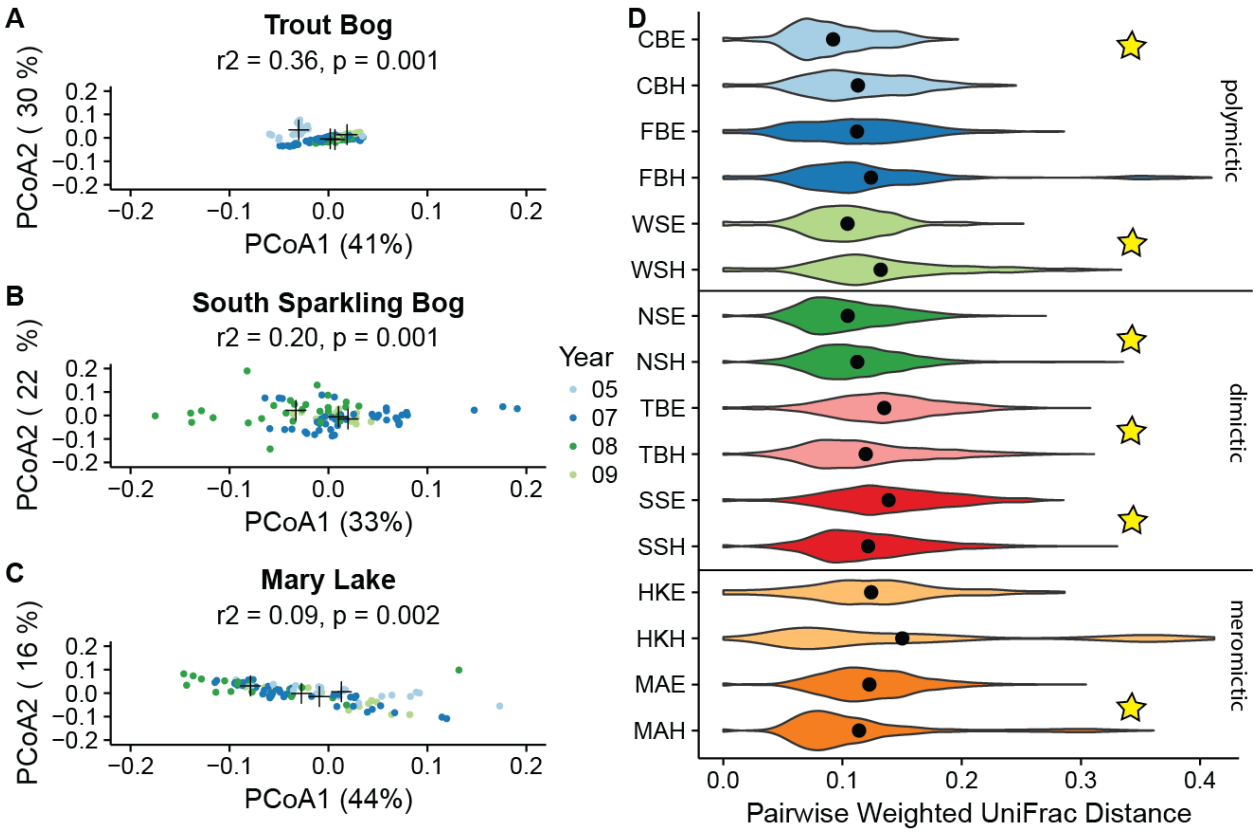
Figure 1. Richness by layer and lake. Lakes on the x axis are arranged by depth (see Table 1 for lake abbreviations and depth measurements). Significance (represented by letters, key below lake IDS in Panel A) was tested using a Wilcoxon signed rank test with a Bonferroni correction for multiple pairwise comparisons, reported in Table S1.



673

674 **Figure 2. Principle coordinates analysis of samples by layer.** Weighted UniFrac distance was
675 used to perform principal coordinates analysis on epilimnion (A) and hypolimnion (B) samples.
676 The percent of variance explained by the first two axes is reported in the axis labels. In both layers,
677 samples cluster significantly by lake and mixing regime as tested using PERMANOVA (Table
678 S2). (See Table 1 for lake abbreviations). Ellipses indicating the clustering of each lake were
679 calculate based on standard error using a 95% confidence interval. Differences in bacterial
680 community composition between lakes and mixing regimes are more pronounced in hypolimnia
681 than epilimnia. Additional plots of this ordination colored by other factors is located in Figure S5.

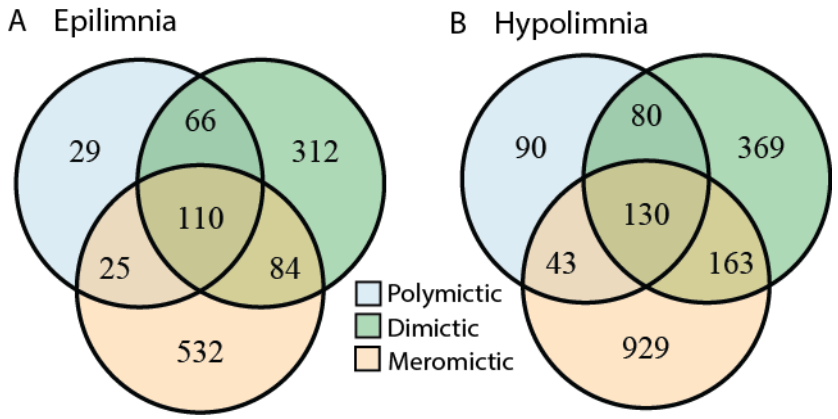
682



683

684 **Figure 3. Internannual variability and dispersion by lake.** Principal coordinates analysis using
685 weighted UniFrac as the distance metric was used to measure the amount of interannual variation
686 in the three lake hypolimnia with the longest time series (A-C). Additional ordinations of lake
687 epilimnia are provided as supplemental figures (Figure S5). Black crosses indicated the centroid
688 for each year. All hypolimnia showed significant clustering by year by PERMANOVA (Table S2).
689 Six outliers in Mary Lake from 2007 are not shown, as their coordinates lie outside the range
690 specified for consistency between plots; these points were included in the PERMANOVA
691 significance test. Panel D shows pairwise weighted UniFrac distance within each lake and layer
692 including all samples. Stars indicate significant differences between layers at $p < 0.05$ by a
693 Wilcoxon signed rank test with a Bonferroni correction for multiple pairwise comparisons.

694



695

696 **Figure 4. Numbers of unique and shared OTUs by mixing regime.** To better understand how
697 shared community membership differs by mixing regime, we quantified the number of shared and
698 unique OTUs in each category. An OTU needed only to appear in one sample at any abundance to
699 be considered present in a category. We found that in both layers, meromictic lakes have the
700 greatest numbers of unique OTUs and polymictic lakes have the least. Meromictic and dimictic
701 lakes shared the most OTUs, while meromictic and polymictic lakes shared the least.

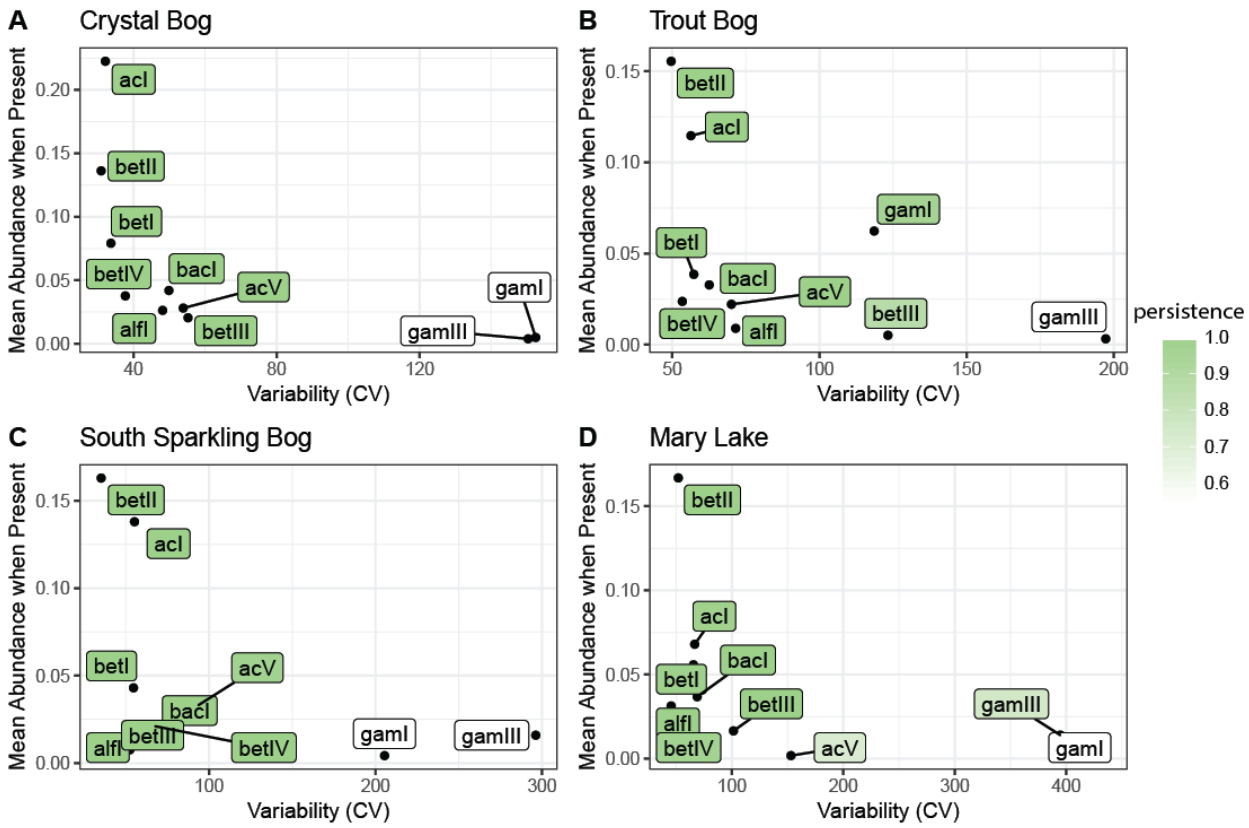


Figure 5. Traits of freshwater lineages. These well-defined freshwater groups showed similar persistence, variance, and abundance in every lake, despite differing abundance patterns. Data from epilimnia with at least two years of undisturbed sampling are shown here. Mean abundance was represented as 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. Additional plots by year can be found in Figure S7.