Freshwater lineages show variable abundance patterns, but consistent traits

Alexandra M. Linz, Benjamin Crary, Ashley Shade, Sarah Owens, Jack A. Gilbert, Rob Knight, and Katherine D. McMahon\*

\* corresponding author: Katherine McMahon, Departments of Civil and Environmental Engineering, and Bacteriology, University of Wisconsin 1550 Linden Drive Room 5552 Madison, WI 53706, Email: trina.mcmahon@wisc.edu

Running title: Freshwater lineages show consistent traits

Keywords:

Abstract word count: 199

Importance word count: 116

Article word count: 4483 (with figure legends)

## Abstract

Bacteria play a key role in freshwater biogeochemical cycling, but controls on long-term freshwater bacterial community composition and dynamics are not well characterized. Here we use a multi-year time series of 16S rRNA gene sequencing data from eight bog lakes to observe annual trends in bacterial communities. Each lake and layer contains a unique bacterial community, with indicator taxa that reflect the environmental conditions of each site. These site differences are likely the result of strong environmental filtering. The community present in each year in each site is also unique, and taxa do not show the same abundance trends each summer. However, each freshwater lineage shows consistent overall persistence, abundance, and coefficient of variance between lakes and years. The high amount of interannual variation observed in this dataset may be explained by unmeasured environmental variables, biotic interactions, or stochastic community assembly. While this introduces new challenges for the prediction of bacterial community dynamics, we found that interannual variation operates within the confines of environmental filtering, and that freshwater lineages have stable traits. The results of our analysis emphasizes the importance of long-term observations, as analyzing only a single year of data would have led us to drastically different conclusions.

## Importance

Lakes are excellent systems for investigating bacterial community dynamics because they have clear boundaries, strong environmental gradients, and well-characterized seasonal changes. The results of our research demonstrate that bacterial community dynamics operate on multi-year timescales, a finding which likely applies to other ecosystems and would have large implications for study design and interpretation. In this specific ecosystem, bog lakes play a disproportionately large role in global carbon cycling, and the information presented here may help refine carbon budgets for these lakes. Finally, all data and code in this study are publically 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

A grand challenge in microbial ecology is to predict bacterial community dynamics. Prediction of change in bacterial communities would be immensely useful in treatment of diseases linked to the human microbiome (1), forecasting cyanobacterial blooms in freshwater (2), and a variety of industrial applications (3–6). However, we have only a cursory knowledge of the factors involved in bacterial community dynamics. Many microbial ecology studies sample large numbers of sites in order to draw conclusions about how bacterial communities are structured, a pursuit which has greatly benefitted our field. But in addition to space, time must also be considered in order to understand change in bacterial communities.

Long-term studies of the bacterial communities present logistical difficulties, but results from the Microbial Observatory projects are promising. For example, the San Pedro North Pacific - Microbial Observatory contributed to our understanding in heterogeneity of bacterial communities (7), and research at the Sapelo Island – Microbial Observatory has led the field in linking genomic data to metadata (8) and the use of environmental RNA (9). In our own North Temperate Lakes – Microbial Observatory, sweeps in diversity at the genome level were detected using metagenomes from several years of sampling (10), enhancing our knowledge of bacterial speciation. Long-term time series are powerful tools for understanding the drivers and controls of bacterial communities.

In this study, we present over 1,300 16S amplicon samples spanning four years, eight lakes, and two depths from the North Temperate Lakes – Microbial Observatory. The sites in this study are bog lakes in the boreal region of northern Wisconsin. Bog lakes are open regions of water surrounded by sphagnum moss, which leaches dissolved organic carbon into the water as of humic and fulvic acids. These ecosystems contain much greater amounts of carbon than of nitrogen or phosphorus and are categorized as “humic” or “dystrophic.” Because the dissolved organic carbon stains the water dark brown, these lakes quickly absorb heat from sunlight, creating strong vertical gradients structured by temperature and dissolved oxygen concentrations (11, 12). Bog lakes mix at different frequencies depending on depth, surface area, and bathymetry (13). These frequencies are called “mixing regime,” and our lakes are classified as either polymictic (more than two mixing events per year), dimictic (exactly two mixing events per year), or meromictic (no mixing events).

Their well-defined boundaries and environmental gradients make bog lakes excellent study systems for microbial ecology, but bog lakes are important ecosystems in their own right. Bogs and wetlands play a crucial role in the global carbon cycle by releasing and consuming methane and carbon dioxide, functions largely performed by microbes (14). The amount of carbon stored and emitted by wetlands is significant on a global scale (15). Understanding bacterial community dynamics would improve our understanding of carbon cycling in freshwater, and therefore our estimates of the global carbon budget.

We use our 16S dataset to analyze bacterial community composition over space and time. We find evidence of environmental filtering by site and identify taxa with preferences for specific locations. We see that broad metrics of bacterial communities repeat annually, but abundance patterns of individual taxa do not. Finally, we examine traits in the abundance and presence/absence patterns of taxa in order to determine if these groups behave in a consistent manner each year. Our research demonstrates that freshwater bacterial communities are driven by factors acting on time scales greater than five years, yet community composition is constrained by site specific factors and the traits of the taxa themselves.

## Methods

**Sample Collection**

Water was collected from eight bog lakes during the summers of 2005, 2007, 2008 and 2009, as previously described (16). Briefly, the epilimnion and hypolimnion layers were collected separately using an integrated water column sampler. Dissolved oxygen and temperature profiles were measured at the time of collection using a handheld YSI 550A (YSI Inc., Yellow Springs, OH). After transport to UW Trout Lake Station, approximately 150 mL from each well-mixed sample was filtered through a 0.22 micron polyethersulfone filter (Supor 200, Pall, Port Washington, NY). Filters were stored at -80C until DNA extraction using the FastDNA Spin Kit for Soil (MP Biomedicals, Santa Ana, CA), with minor modifications (17). The sampling sites are located near Boulder Junction, WI, and were chosen to include lakes represent all three mixing regimes. (Table 1). Trout Bog and Crystal Bog are also primary study sites for the North Temperate Lakes - Long Term Ecological Research Program, which measures a suite of chemical limnology parameters fortnightly during the open water season. The NTL-LTER also maintains autonomous sensing buoys on these lakes, allowing for refined mixing event detection based on thermistor chain measurements

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 *(m2)* | 1300 | 5600 | 4700 | 11900 | 10100 | 4400 | 30000 | 12000 |
| Mixing regime | Polymictic | Polymictic | Dimictic | Polymictic | Dimictic | Dimictic | Meromictic | Meromictic |
| GPS coordinates | 46.05,  -89.65 | 46.01,  -89.61 | 46.00,  -89.71 | 46.00, -  89.71 | 46.04,  -89.69 | 46.04,  -89.71 | 46.19,  -89.70 | 46.25,  -89.90 |
| Years sampled | 2007 | 2007, 2009 | 2007, 2008, 2009 | 2007 | 2005, 2007, 2008, 2009 | 2007, 2008, 2009 | 2007 | 2005, 2007, 2008, 2009 |
| Dissolved inorganic carbon *(ppm)* |  | 0.69, 1.72 | 1.12, 2.31 |  | 1.73, 4.47 | 1.97, 6.42 |  | 5.54, 12.38 |
| Dissolved organic carbon *(ppm)* |  | 15.47, 13.6 | 10.05, 10.40 |  | 19.87, 20.58 | 12.40, 21.92 |  | 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 |

**Table 1. Location and characteristics of bog lakes included in this study.** All lakes are located near Boulder Junction, Wisconsin USA. The mixing regime indicates the frequency of mixing of the two stratified layers, the epilimnion and hypolimnion. Nutrient concentrations are reported for the epilimnion first, then the hypolimnion, and represent the average concentration measured in 2008.

**Sequencing**

1,510 DNA samples, including 547 biological replicates, were sequenced by the Earth Microbiome Project according to their standard protocols (18). 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 (<https://github.com/biocore/deblur>, manuscript in preparation). 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 (19). Thirty-three chimeras and 340 chloroplast sequences (based on pre-clustering and classification with the Greengenes 16S database, May 2013) (20) were removed. Samples were rarefied to 2,500 reads; samples with less than 2,500 reads were omitted, resulting in 1,387 remaining samples. A rarefaction cutoff of 2,500 reads was chosen based on simulations in order to maximize the number of samples retained, while maintaining sufficient quality for downstream analysis of diversity metrics. After these steps, 8,795 OTUs remained.

Each sequence was classified in either our curated freshwater database (21) or the Greengenes database based on the output of NCBI-BLAST (blast+ 2.2.3.1) (22). 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 (23), 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 v3.2.1 (R Development Core Team (2008). R: A language and environment for statistical computing.). Similarity between samples was compared using UniFrac distances, as implement in “phyloseq” (24) (P.J. McMurdie and S. Holmes (2013). phyloseq: An R Package for reproducible interactive analysis and graphic of microbiome census data). Weighted and unweighted Unifrac distance was compared with Bray-Curtis Dissimilarity and Jaccard Similarity, implemented in “vegan” (J. Oksanen, (2016). vegan: Community Ecology Package). Weighted UniFrac distances were chosen for principle coordinates analysis, performed by betadisper() in “vegan”, because it explained the greatest amount of variation in the first two axes. Significant clustering by year in PCoA was tested using PERMANOVA with the function adonis() in “vegan.”

Indicator species analysis was performed using “indicspecies” (25). Only taxa with read abundances in the top 25% of the entire dataset were used for this analysis. The group-normalized coefficient of correlation was chosen for this analysis because it measures both positive and negative habitat preferences and accounts for differences in the number of samples from each site. All taxonomic levels were included in this analysis at the same time to determine which level of resolution was the best indicator for each taxonomic group.

Plots were generated using “ggplot2” (Wickham. (2009). ggplot2: Elegant Graphics for Data Analysis). “reshape2” was used for data formatting (H. Wickham (2007). Reshaping Data with the reshape Package). Data and scripts from this study can be downloaded from the R package “OTUtable” and the McMahon Lab GitHub repository (<https://github.com/McMahonLab/North_Temperate_Lakes-Microbial_Observatory>).

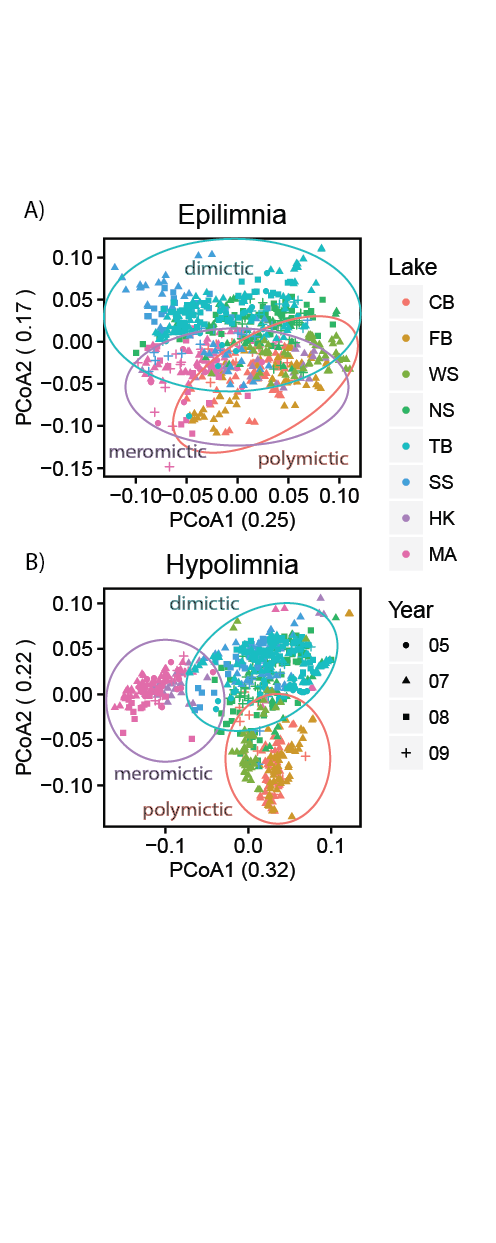
## Results

**Overview of community composition**

A multi-year time series of 16S data from multiple lakes was used to investigate bacterial community dynamics over long time scales. 8,795 OTUs were detected in 1,387 samples. In this time series, Proteobacteria, Actinobacteria, Bacteroidetes, and Verrucomicrobiawere the most abundant phyla. Within these phyla, OTU abundance is highly uneven. For example, much of the abundance of Proteobacteria can be attributed to a few OTUs belonging to the well-known freshwater taxa *Polynucleobacter* and *Limnohabitans,* and the ubiquitous freshwater clade acI contributes much of the observed abundance of Actinobacteria. Unevenness is a recurring theme in this dataset, which has a long rare tail of OTUs and trends driven largely by the most abundant OTUs. Communities from the epilimnion and hypolimnion layers are distinct from each other in all lakes (Figure S3). Epilimnia have a higher proportion of OTUs classified using the established freshwater taxonomy, while hypolimnion communities have more OTUs that cannot be classified past the family level, including several candidate phyla. These differences likely reflect the bias in freshwater microbial ecology towards the surface waters of lakes. Richness also varies by lake, with deeper lakes having more taxa (Figure S1).

**Clusters of community composition**

Differences in community composition are quantified using weighted UniFrac distance. We found mixing regime to be a driver of community composition (Figure 1). This effect is stronger in hypolimnia, which experience major changes in oxygen content during mixing events. One exception is that Crystal Bog, a polymictic lake, clusters with other polymictic hypolimnia in 2007 but with dimictic hypolimnia in 2009. High resolution buoy data for Crystal Bog from the North Temperate Lakes Long Term Ecological Research Project shows that Crystal Bog mixed multiple times in 2007, but only twice in 2009. Within each lake and layer, years have unique and distinct communities. While the change in Crystal Bog can be attributed to mixing frequency, the reasons for this trend in other lakes is less clear. Some possible drivers could include climatic factors, landscape level events, differences in geography, or stochastic community assembly.



**Figure 1. Principal components analysis based on weighted UniFrac distance of samples split by layer.** Clustering by lake, mixing regime, and year are observed are observed in both layers, although the effect is stronger in hypolimnia (r2 = 0.34 by lake and 0.20 by regime in epilimnia, r2 = 0.49 by lake and 0.22 in hypolimnia, all significant at p < 0.05 using PERMANOVA). This suggests that environmental filtering is occurring based on parameters unique to each site.

**Indicator taxa**

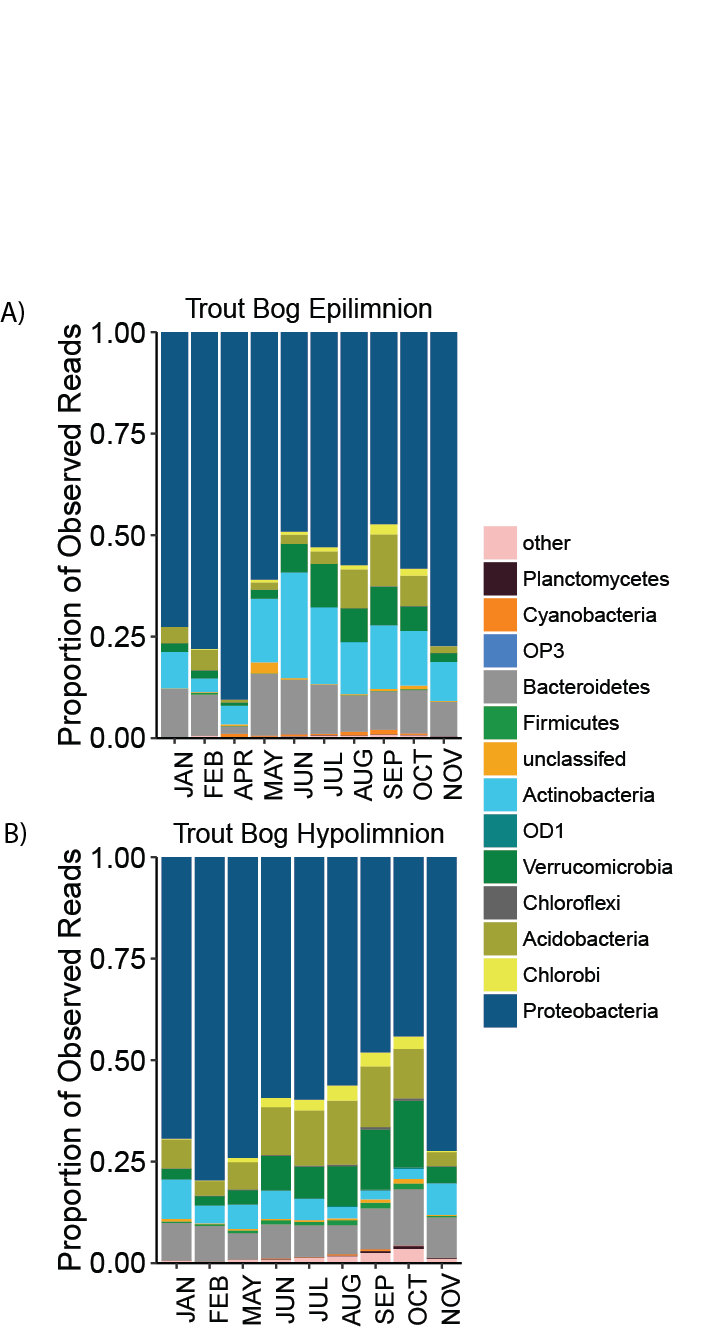
We performed indicator analysis to investigate the taxa driving differences in community composition between mixing regime (Table 2). This technique identifies taxa that are found significantly more often in one group of samples than another; in this case, the groups were defined by layer and mixing regime. There is substantial overlap between the indicator taxa of polymictic epilimnia and hypolimnia, consistent with high mixing frequency. The presence of taxa endemic to each mixing regime likely reflects the environmental filtering and biogeochemical cycling taking place in these ecosystems.

**Seasonal trends**

Despite distinct community compositions in each year of sampling, there are high level seasonal trends. Both biodiversity and evenness increase over time while stratification is in place (Figure S2). At the phylum level, the abundance of Proteobacteriais particularly high in winter and during mixing events (Figure 2). After spring stratification, the abundance of Proteobacteriadecreases as the abundances of other phyla increase. Because this is relative abundance data, this trend may reflect the observed increase in biodiversity over time, and the survivability of dominant Proteobacterial groups *Polynucleobacter* and *Limnohabitans* during the ice-on period of the year. This trend is found in other dimictic lakes as well (Figure S4).

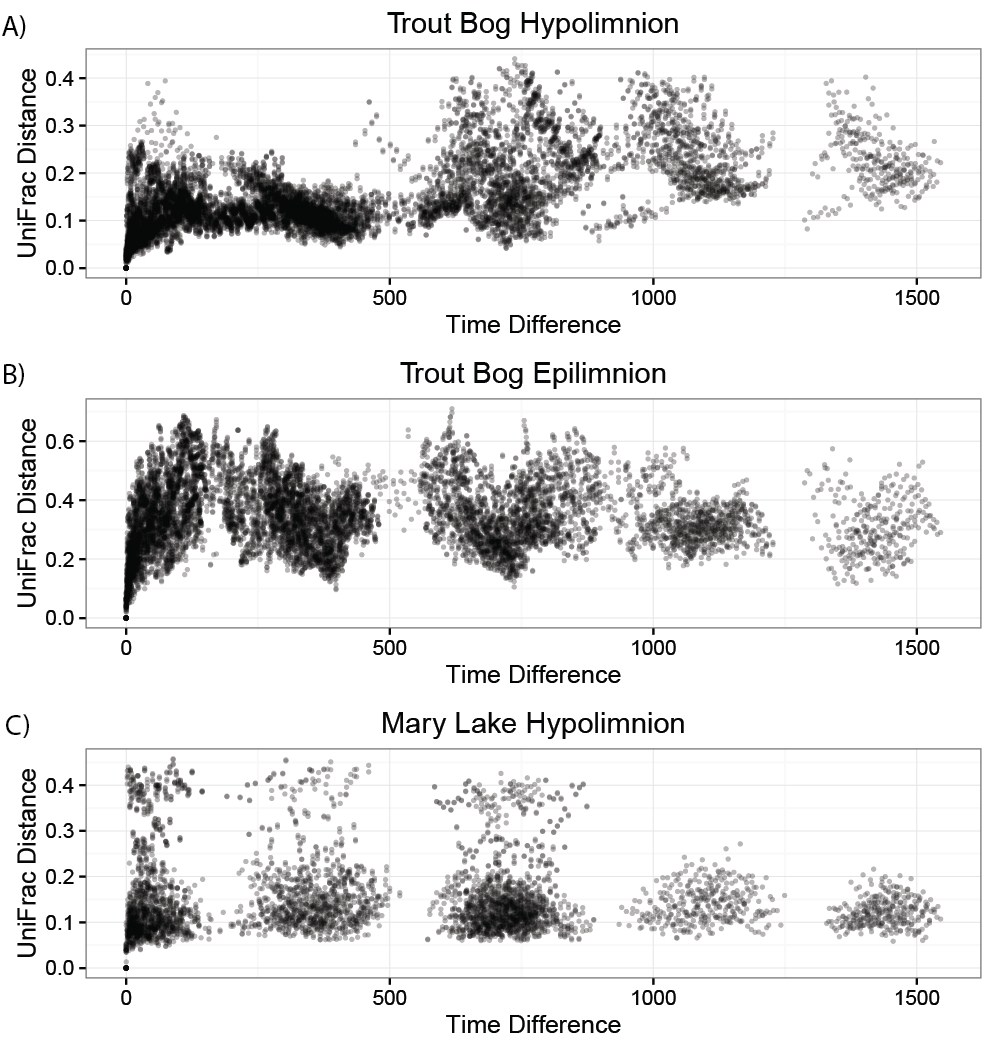
**Table 2. Indicator taxa of layers and mixing regimes.** Indicator analysis was used to identify taxa with a preference for sites defined by layer and mixing regime. OTUs were grouped into higher taxonomic levels, and all levels were used for this analysis at once. The lowest classification of each indicator taxa is reported below.

|  |  |
| --- | --- |
| **Polymictic epilimnia** | **Polymictic hypolimnia** |
| 1. Actinobacteria;Actinobacteria;Actinomycetales;acI;acI-B;acI-B3;OTU2 2. Verrucomicrobia;Spartobacteria; Chthoniobacterales;verI-A 3. Proteobacteria;Betaproteobacteria;Burkholderiales;betI;betI-A;Lhab-A4;OTU2 4. Proteobacteria;Betaproteobacteria;Methylophilales;OTU4 5. Actinobacteria;Actinobacteria;Actinomycetales;acI;acI-B;acI-B3;OTU6 6. Proteobacteria;Betaproteobacteria;Burkholderiales;betI;betI-A 7. Verrucomicrobia;Spartobacteria 8. Actinobacteria;Actinobacteria;ActinomycetalesMycobacteriaceae;Mycobacterium;OTU2 9. Actinobacteria;Actinobacteria;Actinomycetales;acI 10. Proteobacteria;Betaproteobacteria;Burkholderiales;betII;Pnec;PnecC | 1. Actinobacteria;Actinobacteria;Actinomycetales 2. Actinobacteria 3. Proteobacteria;\_Betaproteobacteria;Burkholderiales;betI;betI-A 4. Verrucomicrobia;Spartobacteria;Chthoniobacterales;verI-A 5. Bacteroidetes;Saprospirae 6. Proteobacteria;Alphaproteobacteria;Rhizobiales 7. Proteobacteria;Betaproteobacteria;Burkholderiales;betII;Pnec;PnecC 8. Actinobacteria;Acidimicrobiia;Acidimicrobiales;OTU2 9. Actinobacteria;Acidimicrobiia;Acidimicrobiales;OTU20 10. Proteobacteria;Alphaproteobacteria |
| **Dimictic epilimnia** | **Dimictic hypolimnia** |
| 1. Proteobacteria;Betaproteobacteria;Methylophilales 2. Actinobacteria;Acidimicrobiia;Acidimicrobiales;acV 3. Actinobacteria;Actinobacteria;Actinomycetales;acI;acI-B;acI-B2;OTU3 4. Bacteroidetes;Saprospirae;Saprospirales;Chitinophagaceae;OTU32 5. Bacteroidetes;Sphingobacteriia;Sphingobacteriales;bacVI;bacVI-A;Muci;OTU3 6. Proteobacteria;Betaproteobacteria;Burkholderiales;betI;betI-B;Rhodo;OTU3 7. Acidobacteria 8. Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;OTU2 9. Chlorobi;Chlorobia 10. Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;OTU26 | 1. Acidobacteria;Holophagae;Holophagales;Holophagaceae;OTU2 2. Acidobacteria 3. Proteobacteria;Betaproteobacteria;Burkholderiales;betII;Pnec;PnecC;OTU3 4. Proteobacteria;Betaproteobacteria;Methylophilales;OTU5 5. Verrucomicrobia;Pedosphaerae;Pedosphaerales;Ellin515;OTU2 6. Proteobacteria;Betaproteobacteria 7. Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae;Desulfobulbus 8. Proteobacteria;Betaproteobacteria;Burkholderiales;betI;betI-B;Rhodo;OTU3 9. Proteobacteria;Betaproteobacteria;Methylophilales;Methylophilaceae 10. Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae |
| **Meromictic epilimnia** | **Meromictic hypolimnia** |
| 1. Planctomycetes 2. Actinobacteria;Actinobacteria;Actinomycetales;acTH2 3. Proteobacteria;Betaproteobacteria;Burkholderiales;betI;betI-A;Lhab-A11 4. Actinobacteria;Actinobacteria;Actinomycetales;acI;acI-A;acI-A6 5. Proteobacteria;Alphaproteobacteria;Rhizobiales;alfI;alfI-B;alfI-B21 6. Proteobacteria;Betaproteobacteria;Burkholderiales;betIII;betIII-A;betIII-A1 7. Actinobacteria;Actinobacteria;Actinomycetales;acI;acI-A 8. Proteobacteria;Gammaproteobacteria;Xanthomonadales;Sinobacteraceae;OTU13 9. Proteobacteria;Alphaproteobacteria;Rickettsiales 10. Proteobacteria;Betaproteobacteria;Burkholderiales;betI;betI-A;Lhab-A1;OTU2 | 1. Proteobacteria;Deltaproteobacteria;Syntrophobacterales 2. Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophaceae 3. Omnitrophica;BD4-9 4. Omintrophica 5. Verrucomicrobia;Verruco-5;WCHB1-41;OTU2 6. Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobacteraceae 7. Planctomycetes 8. Archaea;Parvarchaeota 9. Verrucomicrobia;Verruco-5 10. Aminicenantes;OP8\_1 |

****

**Figure 2. Phylum-level seasonal trends.** The proportion of Proteobacteria is highest in winter, early spring, and fall, but decreases during the stratified summer months. Most other phyla show the opposite trend. Because this is based on relative abundance data, this trend likely reflects the observed increase in biodiversity with time, and adds that members of Proteobacteria survive better than other taxa during mixing events and ice coverage.

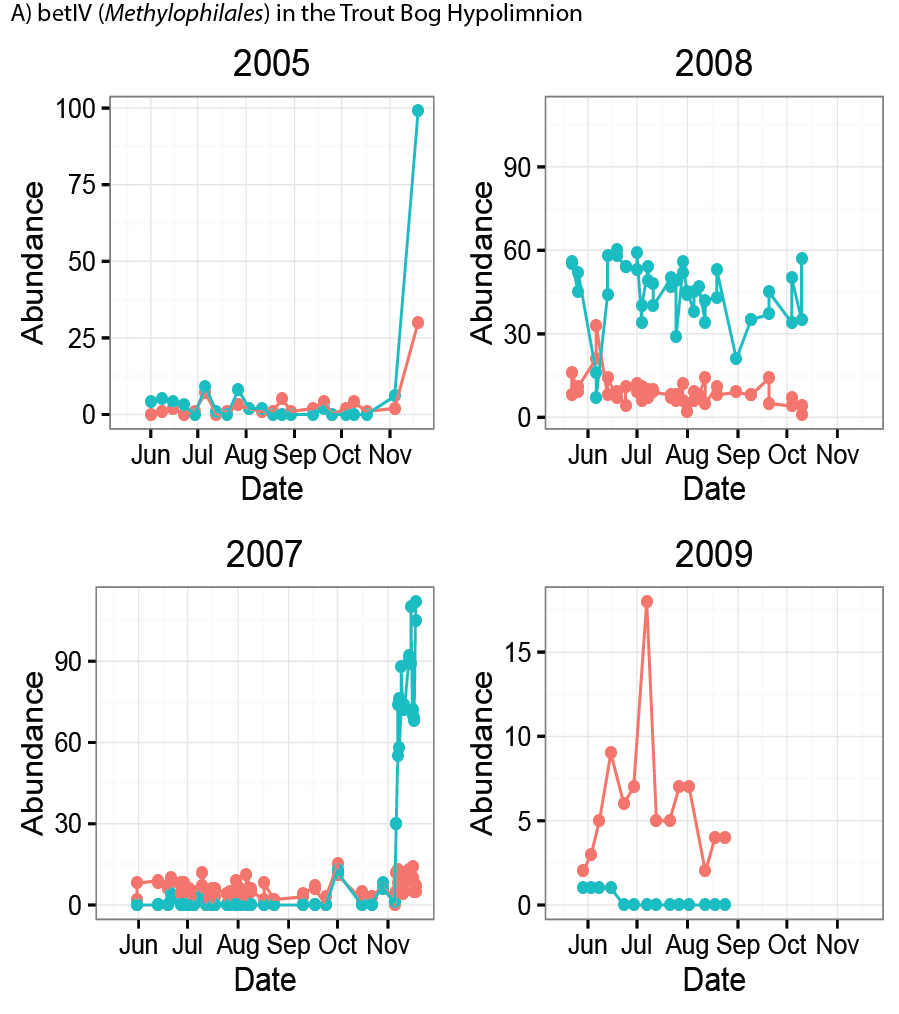
However, seasonal trends are not observed in higher resolution taxonomic levels. We analyzed seasonal trends by plotting UniFrac distance between samples versus the time between sample collection (Figure 3). Based on previous marine studies (26, 27), this was expected to produce a sine wave-like pattern with decreasing amplitude and peaks approximately 365 days apart. When we perform this analysis, we do not see that pattern. Instead, we see similarity decreasing over time in a logarithmic fashion, so that samples close together in time decrease in similarity quickly and that samples taken a year apart are as dissimilar as samples taken four years apart. The only instance of a sine wave-like pattern is observed in Trout Bog, which has increases in similarity at approximately one year and two years. This pattern is driven by the higher number of fall, winter, and spring samples collected from Trout Bog than from other lakes. Samples from these times of year are more similar from year to year, while samples collected during the summer stratification period do not show annual trends at the OTU level. Additional plots from other sites can be found in Figure S3.

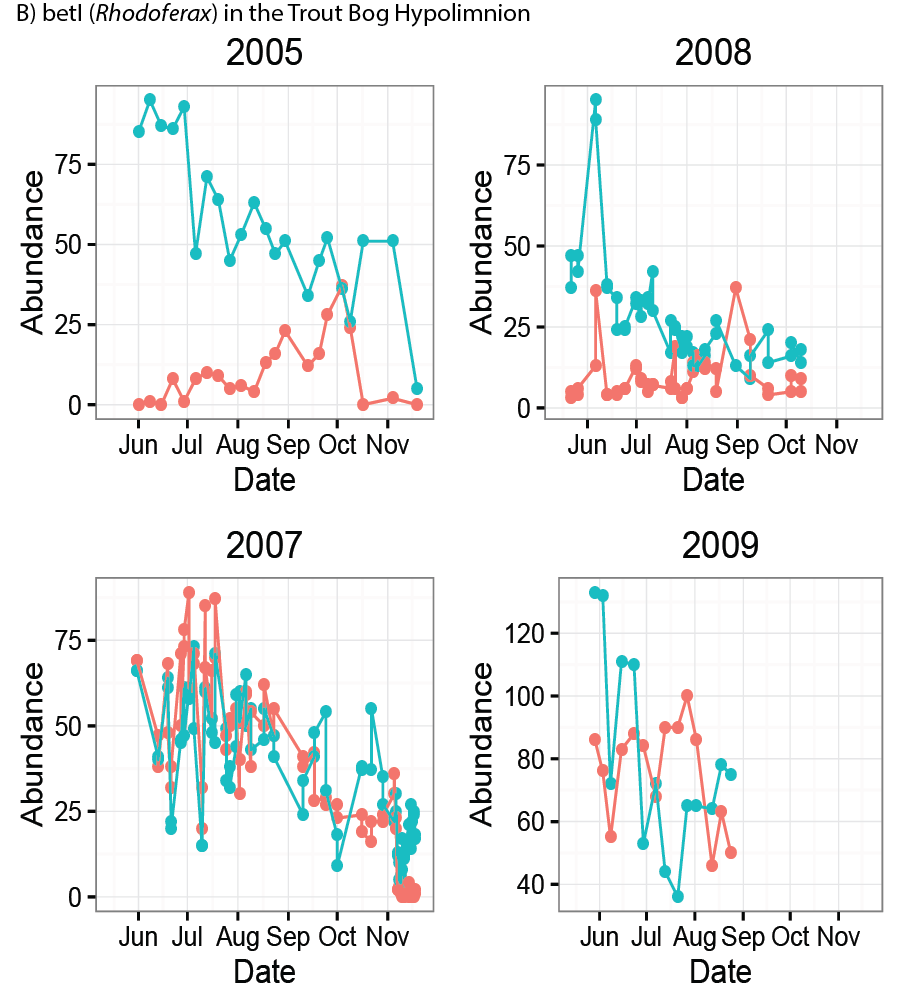


**Figure 3. Lack of seasonal trends at the OTU level.** Despite strong seasonal climatic factors in the studied systems, repeatable seasonal trends were not observed in community composition at the phylum level. The only increased similarity on annual scales was observed in Trout Bog, where high numbers of fall, winter, and spring samples contribute to this trend.

**Trends in specific OTUs do not repeat annually**

As suggested by the lack of similarity between samples taken one year apart, trends in OTU abundance over time do not repeat annually (Figure 4). They do not peak in abundance at consistent times of year or show the same relationship to time each year. OTUs would likely show a consistent response to mixing events because this is such a large disturbance; however, this dataset does not capture a sufficient number of mixing events to demonstrate this. To further complicate OTU abundances, closely related OTUs have different interactions in different years. We have observed the same OTUs to be strongly correlated in some years while strongly anti-correlated in other years. Additional examples can be found in Figure S8.

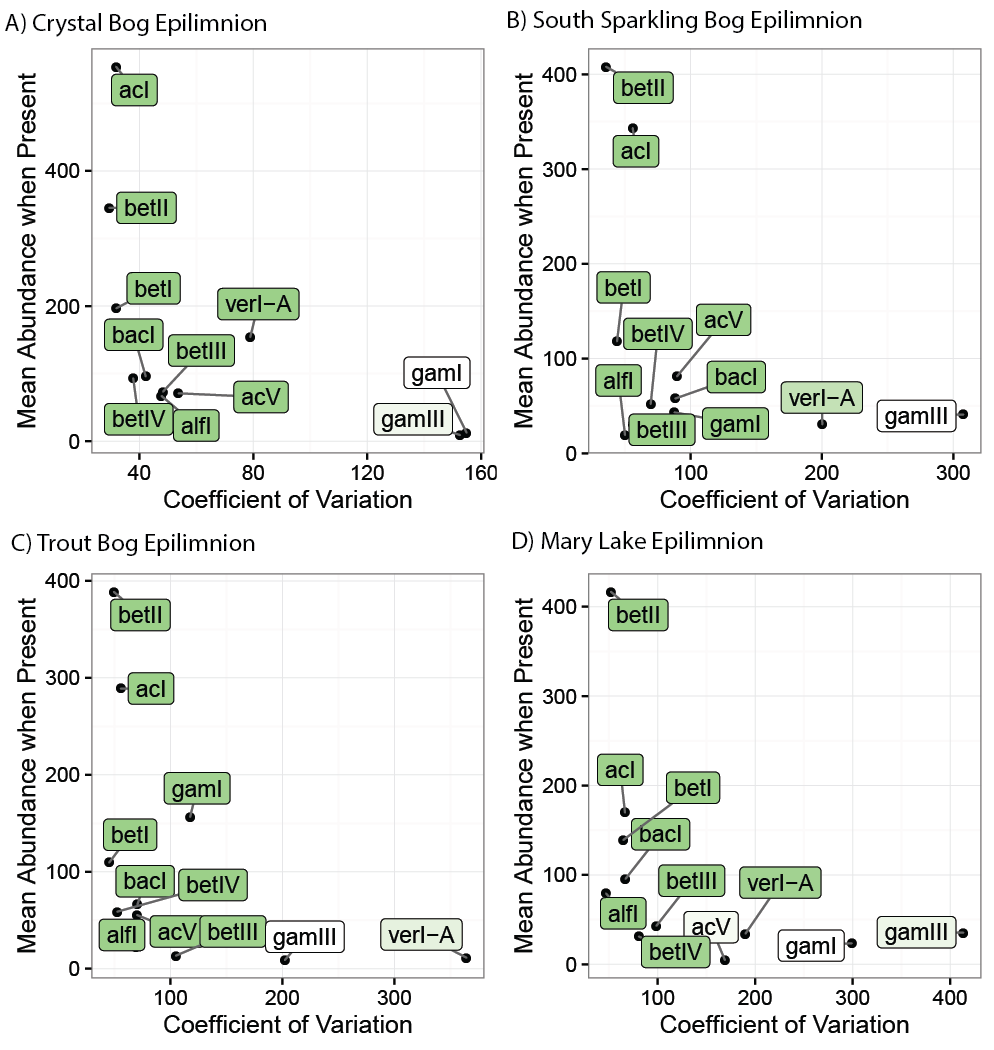
****

****

**Figure 4. Examples of different trends in abundance with closely related OTUs.** Panel 1 shows clades of *Methylophilales* in the Trout Bog Hypolimnion. While 2005 and 2007 both show low abundance until a sharp increase in November, 2008 and 2009 show more linear abundance patterns, with a different clade dominant in each year. In Panel 2, two OTUs classified in the Rhodo clade are shown. The OTUs are positively correlated in 2007, negatively correlated in 2005, and have no strong correlation in either direction in 2008 and 2009. Examples of more clades and lakes can be found in the supplemental document.

**Consistent traits of OTUs**

Even though OTUs do not show the same trends each year, they do have traits that are consistent between years and lakes. We quantified mean abundance when present, persistence, and the coefficient of variance (CV) for lineages in the freshwater taxonomy, metrics which have been previously used to categorize OTUs (28, 29) (Figure 5). OTUs classified in the freshwater taxonomy were summed into lineages for this analysis. We learned that low persistence is associated with a high CV and that a low CV is associated with high abundance. We rarely observe “bloomers,” situations where a clade has both high abundance and low persistence. Most freshwater lineages are highly persistent at low abundances with a low CV. Clade gamIII of the Gammaproteobacteria is an exception, with low persistence, low abundance, and a high CV. Clades gamI and verI-A occasionally also exhibit this profile. Clades betII and acI are highly abundant and persistent with a low CV, consistent with their profile as ubiquitous freshwater generalists. These traits hold from year to year in the same lake (Figure S6). Knowledge of the general manner in which these clades behave can begin to shed light on their lifestyles and the drivers of their observed abundance trends.



**Figure 5. Traits of freshwater clades.** These well-defined freshwater clades show similar persistence, coefficient of variance, and abundance in each year and lake, despite different abundance patterns. This suggests that unknown functions or other metabolic characteristics are driving a stable lifestyle.

## Discussion

After viewing the lack of repeatable annual trends in Figure 4, it is tempting to throw one’s hands in the air and declare any attempt to predict freshwater bacterial community composition a lost cause. We urge readers not to do so just yet. Despite high interannual variation, freshwater lineages themselves have predictable traits. We see that persistence, abundance, and coefficient of variance of freshwater groups are relatively consistent between years and lakes. But if this is the case, why are there such large differences between years?

Based on the distinct community compositions between epilimnia and hypolimnia, and between lakes, environmental filtering likely has a strong effect in bog lakes. This is consistent with previous research on these bog lakes (16) and supported by increased phylogenetic clustering (30) (supplemental). The identities of indicator taxa for mixing regimes reflect the biogeochemical properties of their sites. For example, Desulfobulbaceae and Syntrophobacterales are indicators for the anoxic, sulfur-rich samples from dimictic and meromictic hypolimnia, consistent with their probable function of anaerobic sulfur reduction. In contrast, Actinobacteria and particularly its freshwater lineage acI are indicators exclusively for epilimnia and polymictic hypolimnia, where they can likely take full advantage of sunlight to power actinorhodopsins (31). Interestingly, candidate phyla Omnitrophica (OP3) and Aminicenantes (OP8) were found to be indicators of meromictic hypolimnia, consistent with their detection in anoxic freshwater (32, 33). Parcubacteria (OD1), a candidate phylum previously identified in bog lakes (34), was not an indicator of mixing regime as it was found predominantly in a single lake, South Sparkling Bog.

One possible explanation for the observed interannual variation is that there are abiotic drivers not measured that vary strongly from year to year. Changes in the terrestrial ecosystem surrounding the aquatic systems studied may have profound impacts on bacterial community composition. Biotic interactions may also lead to variation. Bacterial communities have been shown to take part in complex networks (35), and forces such as competition and cooperation may be structuring community composition in addition to environmental filtering. This study does not include information on phytoplankton or dinoflagellate abundances, which have been shown to influence bacterial community composition in bog lakes (36, 37). Finally, stochastic community assembly may occur after the disturbances of mixing events and the winter ice-on period, which appear to be bottlenecks reducing diversity in the bacterial community (38). The emergence of distinctive epilimnia and hypolimnia communities after the spring thaw and mixing event may introduce an element of chance in community assembly, resulting in unique communities each year.

There is a large body of work on predictable seasonal trends in oceans, rivers, and freshwater lakes (27, 39–41). One major difference between these sites and ours is that these other systems are larger and contain currents or other water movement, while bog lakes are more stagnant and are connected to other aquatic systems via only groundwater. Another difference is that many of the systems studied do not experience the temperature extremes that boreal bog lakes do. The impact of winter freezing and mixing events may prevent formation of a stable community each year. The study most similar to ours is a seasonal analysis of Lake Mendota (17). Lake Mendota freezes annually and is close to the observed study sites geographically, but is a lot bigger than the bogs, is part of a larger chain of lakes with more water exchange, and doesn’t stratify as quickly. The Mendota study notes that late summer samples are more difficult to predict than fall, winter, or early spring samples. Similarly, a comparison of seasonal trends between Lake Mendota and Crystal Bog found stable bacterial community composition in spring and fall, but “dramatic changes” in composition during summer (42). Since our bog lake dataset is dominated by stratified summer samples, this would make annual prediction far more difficult.

Perhaps the biggest implication of this research is the importance of long-term time series. A similar dataset spanning only a single year would have produced different conclusions about the seasonality of freshwater bacterial communities. Clearly, the factors driving community composition in bog lake ecosystems are more complex than we had originally thought. It is possible that trends may repeat on scales greater than the five years covered in this dataset, or that annual differences are driven by environmental factors that do not occur every year. To answer these questions, we are continuing to collect and sequence samples for the North Temperate Lakes – Microbial Observatory, and we are expanding our sequencing repertoire beyond 16S. All of the 16S data we have currently generated can be found in the R package “OTUtable” which is available on CRAN for installation via the R command line, or on our GitHub page. We hope that this dataset and its future expansion will be used as a resource for researchers investigating their own questions about how bacterial communities behave on long time scales.

**Conclusion**

Using a multi-year, multi-lake time series, we find depth and mixing frequency to be major drivers of community composition, and identify specific bacterial taxa associated with each layer and mixing regime. Trends that are consistent annually are observed in community metrics such as richness and evenness, and in community composition at the phylum level. However, repeatable trends in abundance over time are not identified at higher resolution taxonomic levels. Each year in each lake harbors a unique bacterial community, within the confines of its environmental parameters. While this initially appears to pose a challenge to the prediction of bacterial community composition, we also demonstrate that traits such as persistence, abundance, and coefficient of variance are consistent in freshwater taxa despite varying abundance trends each year. Interannual variation could be due to either environmental factors that operate on long time scales, biotic interactions, or stochastic community assembly as the bacterial community establishes after the spring thaw. Our results emphasize the importance of long term time series in microbial ecology in order to draw accurate conclusions. More research is needed to be able to predict bacterial community composition, but the information we have gathered on traits of bacterial taxa and environmental filtering bring us closer to this ultimate goal.

## Acknowledgements

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 Barvaz Sini 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 training set. We acknowledge efforts by many McMahon Lab undergrads and technicians related to sample collection and DNA extraction, particularly Georgia Wolfe.

## References

1. **Cho I**, **Blaser MJ**. 2012. The human microbiome: at the interface of health and disease. Nat Rev Genet **13**:260–270.

2. **Rastogi RP**, **Madamwar D**, **Incharoensakdi A**. 2015. Bloom Dynamics of Cyanobacteria and Their Toxins: Environmental Health Impacts and Mitigation Strategies. Front Microbiol **6**:1–22.

3. **Logan BE**, **Rabaey K**. 2012. Conversion of Wastes into Bioelectricity and Chemicals by Using Microbial Electrochemical Technologies. Science (80- ) **337**:686–690.

4. **Lopes T**, **Gouveia L**, **Reis A**. 2013. Integrated microbial processes for biofuels and high value-added products : the way to improve the cost effectiveness of biofuel production. Appl Microbiol Biotechnol.

5. **Ferrera I**, **Sánchez O**. 2016. Insights into microbial diversity in wastewater treatment systems: How far have we come? Biotechnol Adv in press.

6. **Bourdichon F**, **Casaregola S**, **Farrokh C**, **Frisvad JC**, **Gerds ML**, **Hammes WP**, **Harnett J**, **Huys G**, **Laulund S**, **Ouwehand A**, **Powell IB**, **Prajapati JB**, **Seto Y**, **Ter E**, **Boven A Van**, **Vankerckhoven V**, **Zgoda A**, **Tuijtelaars S**, **Bech E**. 2012. Food fermentations: Microorganisms with technological beneficial use. Int J Food Microbiol **154**:87–97.

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

8. **Sheldon WM**, **Moran MA**, **Hollibaugh JT**. 2000. Efforts to Link Ecological Metadata with Bacterial Gene Sequences at the Sapelo Island Microbial Observatory.

9. **Hollibaugh JT**, **Gifford SM**, **Moran MA**, **Ross MJ**, **Sharma S**, **Tolar BB**. 2014. Seasonal variation in the metatranscriptomes of a Thaumarchaeota population from SE USA coastal waters. ISME J **8**:685–98.

10. **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 1–13.

11. **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.

12. **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.

13. **Read JS**, **Winslow LA**, **Hansen GJA**, **Van den Hoek J**, **Hanson PC**, **Bruce LC**, **Markfort CD**. 2014. Simulating 2368 temperate lakes reveals weak coherence in stratification phenology. Ecol Modell **291**:142–150.

14. **Riera JL**, **Schindler JE**, **Kratz TK**. 1999. Seasonal dynamics of carbon dioxide and methane in two clear-water lakes and two bog lakes in northern Wisconsin, U.S.A. Can J Fish Aquat Sci **274**:1–10.

15. **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.

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

17. **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.

18. **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.

19. **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.

20. **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.

21. **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.

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

23. **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.

24. **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.

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

26. **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.

27. **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.

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

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

30. **Kembel SW**. 2009. Disentangling niche and neutral influences on community assembly : assessing the performance of community phylogenetic structure tests. Ecol Lett **12**:949–960.

31. **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 n/a–n/a.

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

33. **Glöckner J**, **Kube M**, **Shrestha PM**, **Weber M**, **Glöckner FO**, **Reinhardt R**, **Liesack W**. 2010. Phylogenetic diversity and metagenomics of candidate division OP3. Environ Microbiol **12**:1218–1229.

34. **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 J **6**:1640–1652.

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

36. **Newton RJ**, **Kent AD**, **Triplett EW**, **McMahon KD**. 2006. Microbial community dynamics in a humic lake: Differential persistence of common freshwater phylotypes. Environ Microbiol **8**:956–970.

37. **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.

38. **Roguet A**, **Laigle GS**, **Therial C**, **Bressy A**, **Soulignac F**, **Catherine A**, **Lacroix G**, **Jardillier L**, **Bonhomme C**, **Lerch TZ**, **Lucas FS**. 2015. Neutral community model explains the bacterial community assembly in freshwater lakes. FEMS Microbiol Ecol **91**:1–11.

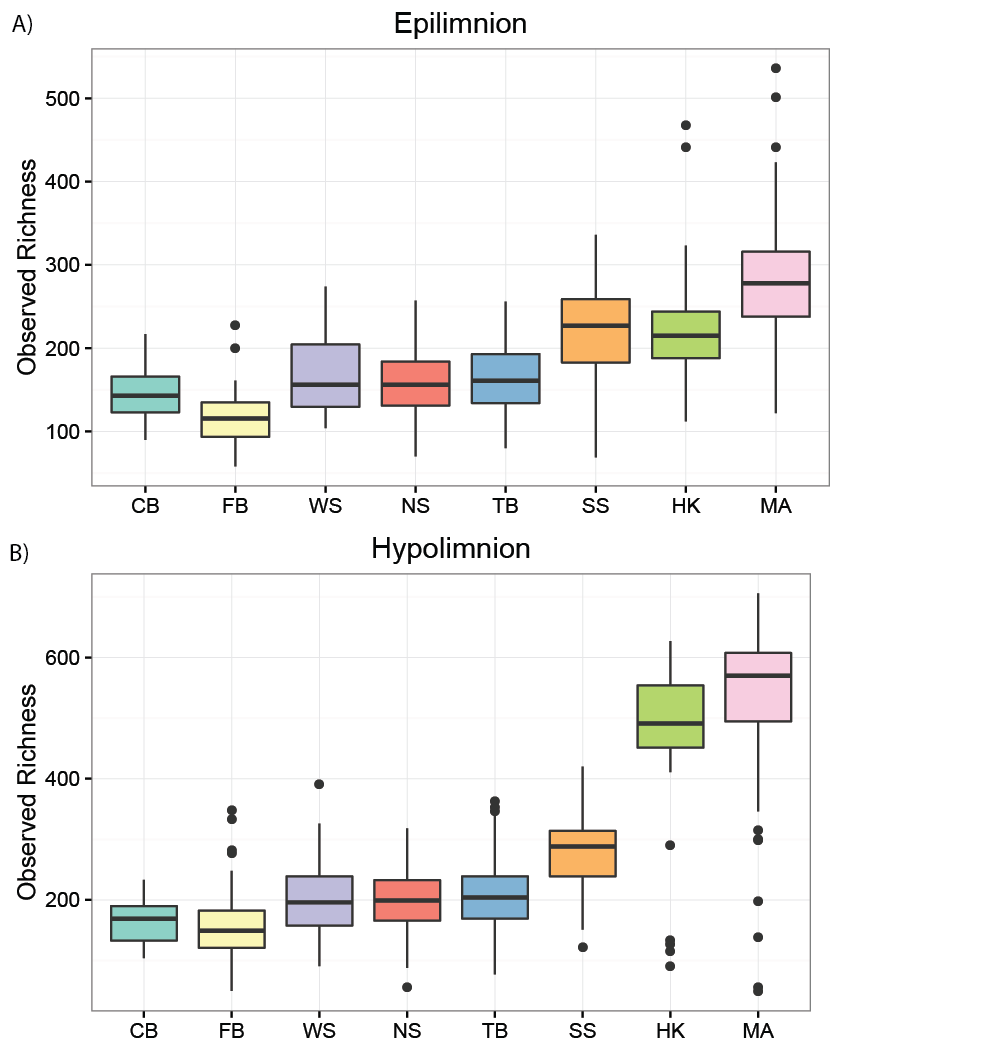
39. **Cram J**, **Chow C**, **Sachdeva R**, **Needham D**, **Parada A**, **Steele J**, **Fuhrman J**. 2015. Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. ISME J **9**:563–580.

40. **Crump BC**, **Peterson BJ**, **Raymond PA**, **Amon RMW**, **Rinehart A**, **McClelland JW**, **Holmes RM**. 2009. Circumpolar synchrony in big river bacterioplankton. Proc Natl Acad Sci USA **106**:21208–12.

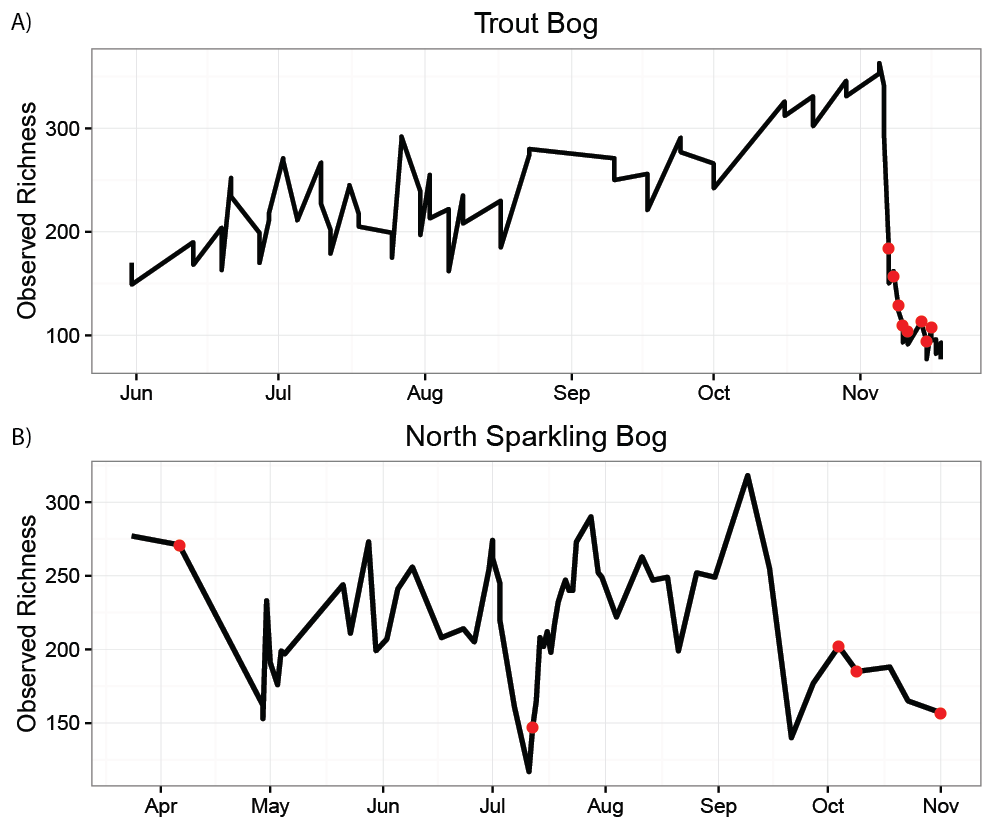
41. **Shade A**, **Chiu CY**, **McMahon KD**. 2010. Seasonal and episodic lake mixing stimulate differential planktonic bacterial dynamics. Microb Ecol **59**:546–554.

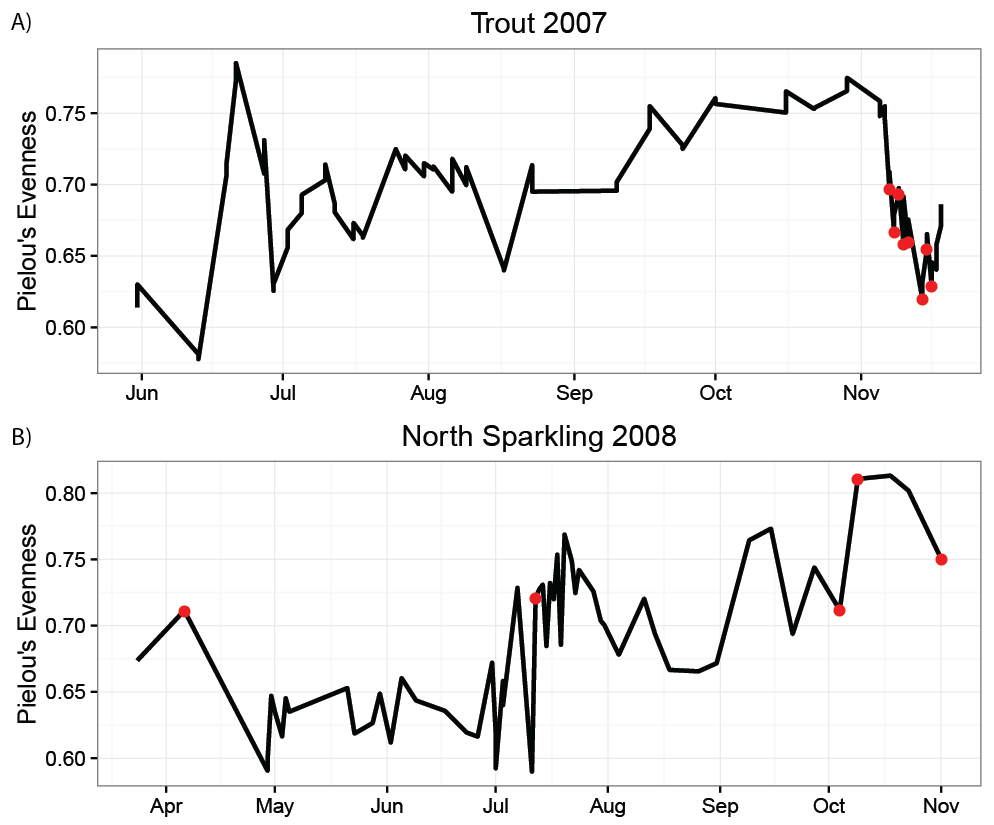
42. **Yannarell AC**, **Kent AD**, **H. LG**, **Kratz TK**, **Triplett EW**. 2003. Temporal Patterns in Bacterial Communities in Three Temperate Lakes of Different Trophic Status. Microb Ecol **46**:391–405.

## Supplemental Figures

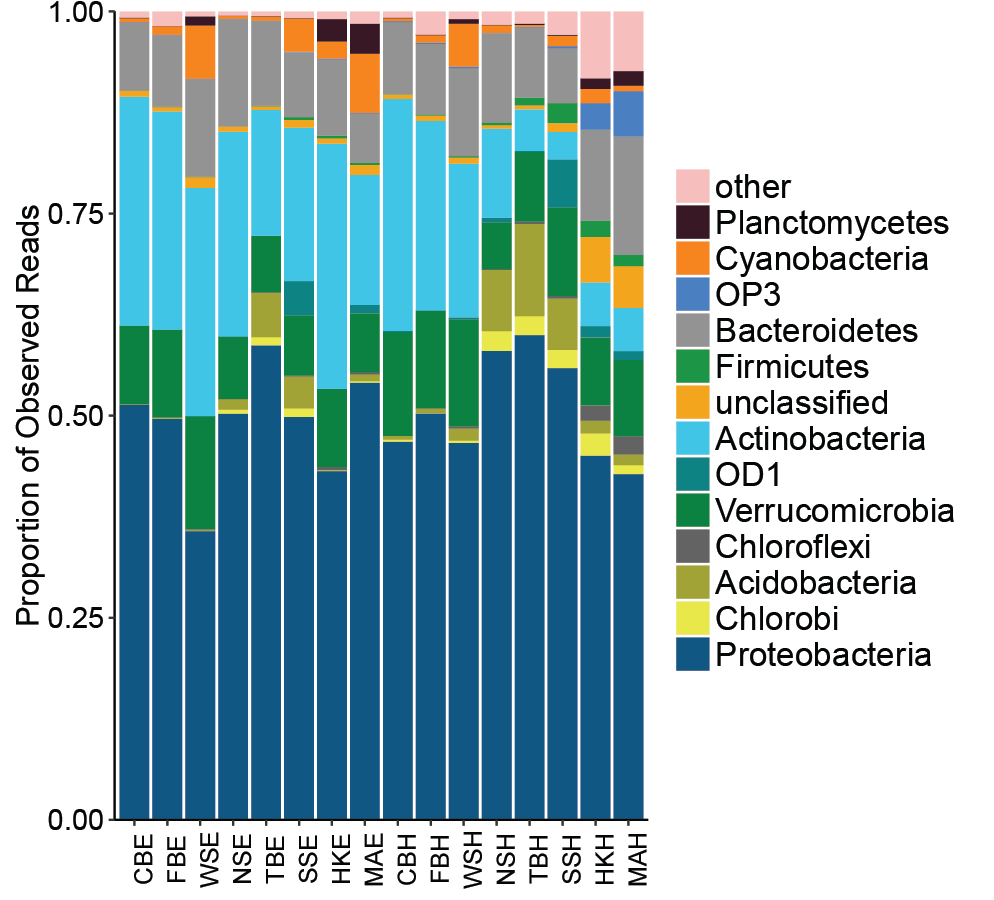


**Figure S1. Observed richness by lake.** Observed richness is correlated with the depth of the lake. This effect is stronger in hypolimnia. Differences in richness between lakes of different mixing regimes are significant at p > 0.05 using a pairwise Wilcoxon Test with a Bonferroni adjustment.

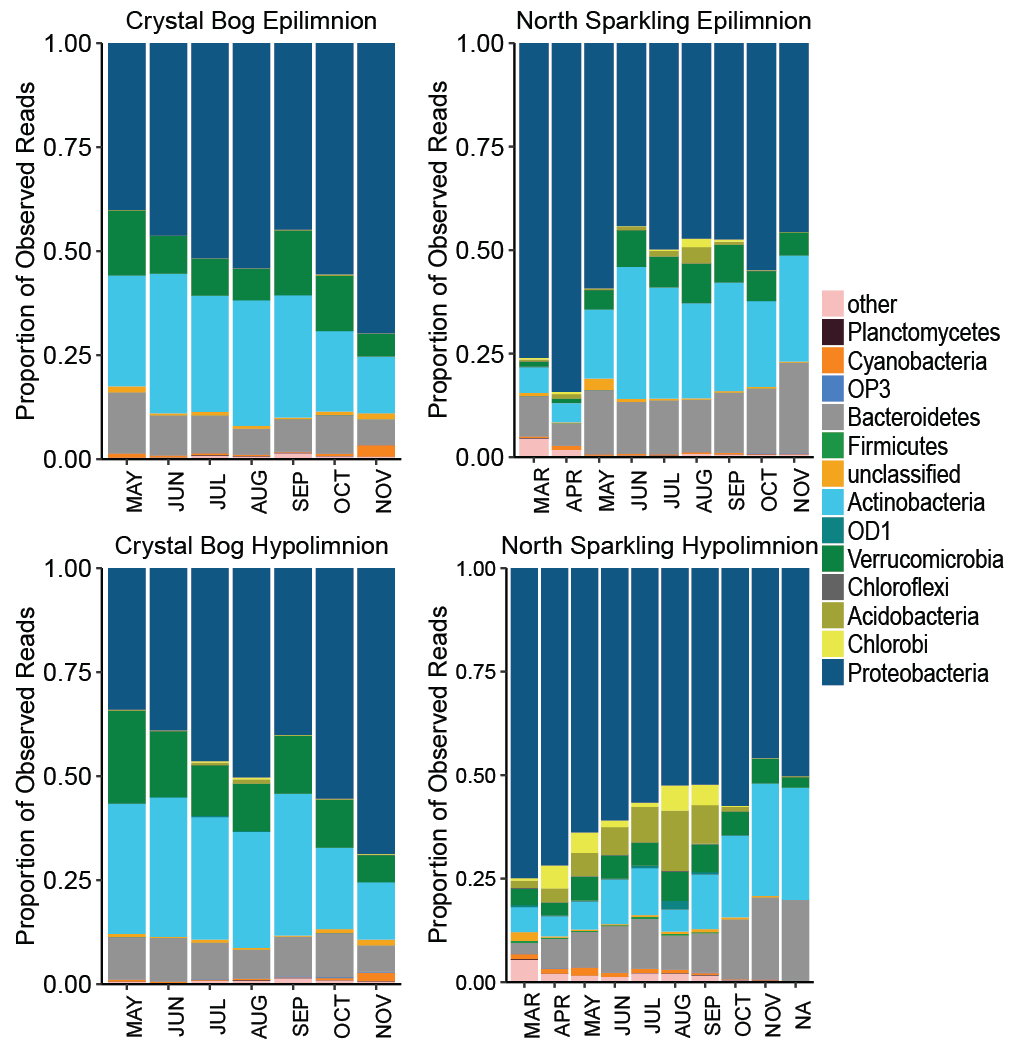


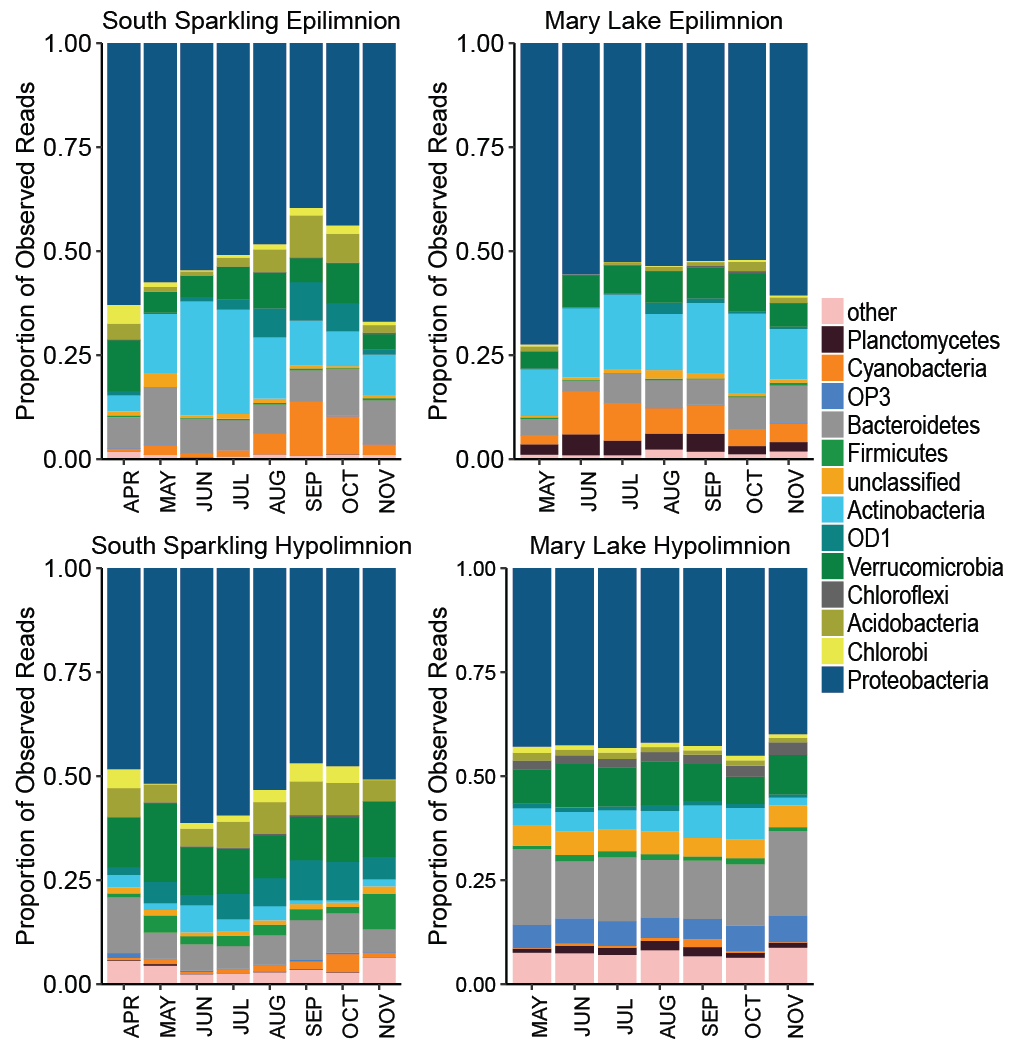


**Figure S2. Observed richness and evenness over time.** Richness tended to decrease during mixing events, marked here in red. In Trout Bog, richness increased steadily during 2007 until the fall mixing event. In North Sparkling Bog, 2008, both the fall mixing event and the artificial mixing event in July show lower richness than the rest of the year.

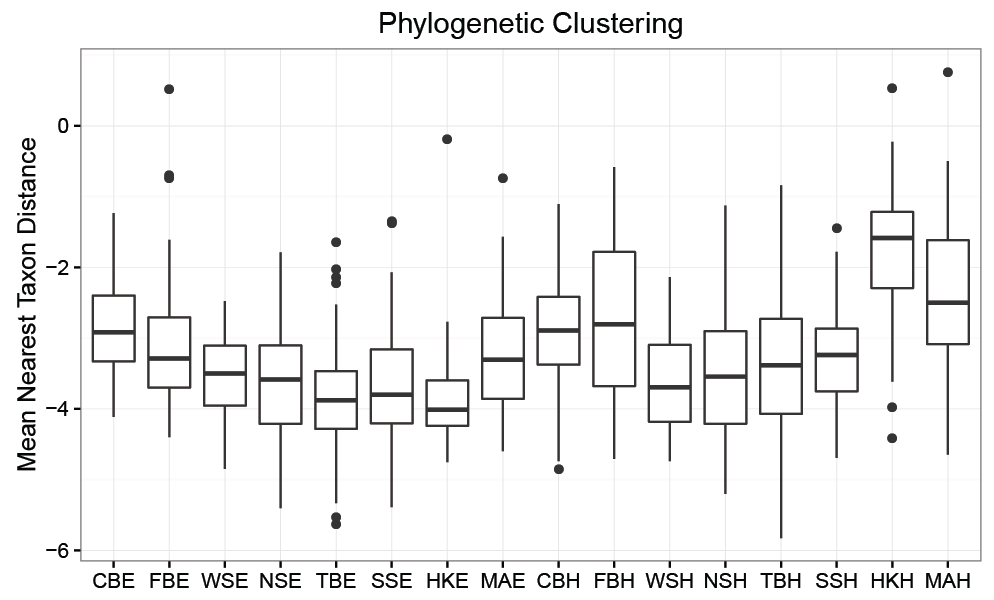
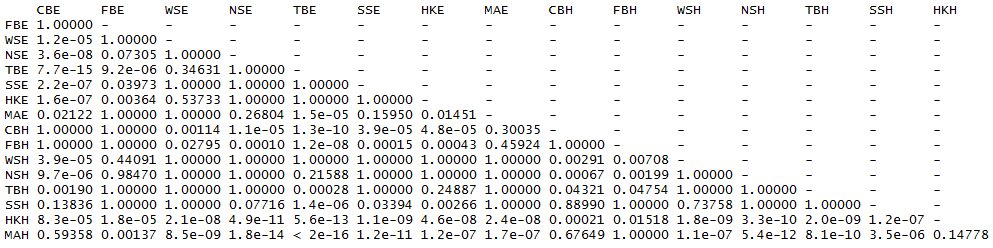


**Figure S3. Phylum composition by lake and layer.** Phylum composition differs by sampling site. A few phyla dominate everywhere, including Proteobacteria, Verrucomicrobia, Actinobacteria, and Bacteroidetes. However, proportions of these phyla decrease in meromicticlakes, which have greater proportions of Planctomyces, OP3, and all lower abundance and other phyla. OD1 is found primarily in South Sparkling Bog, and Acidobacteria are more abundant in dimictic lakes South Sparkling, North Sparkling, and Trout Bog.

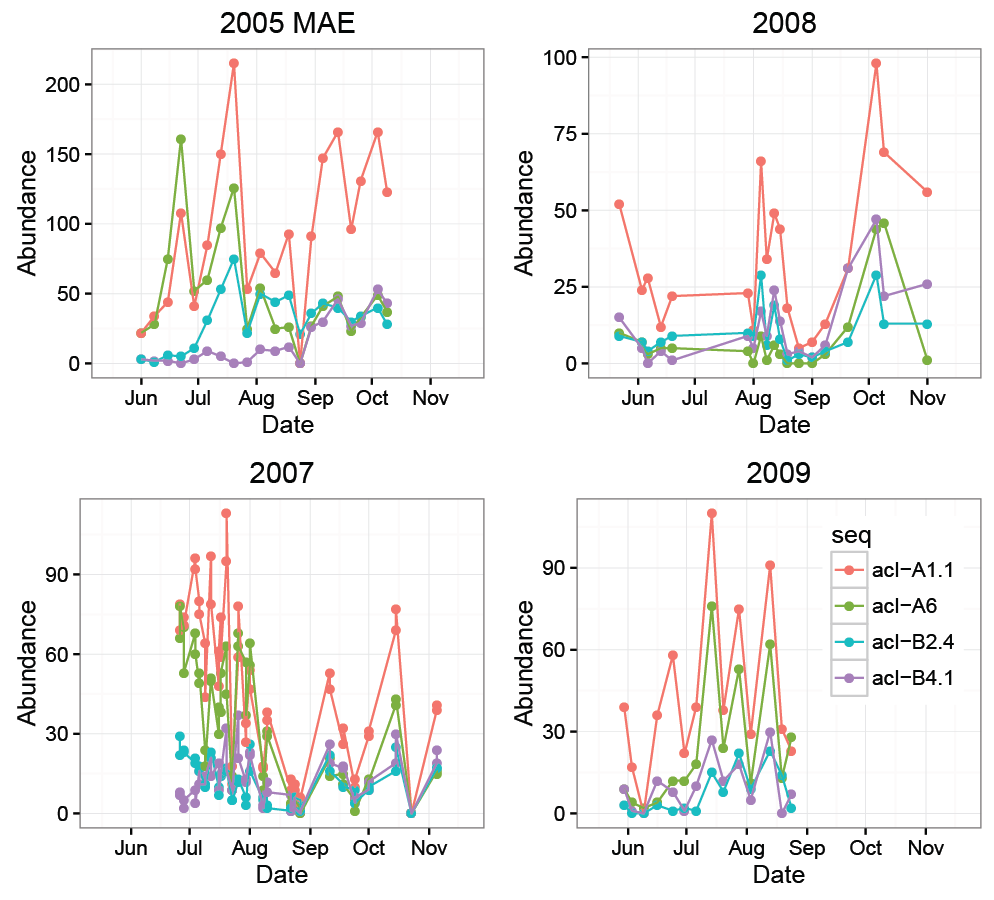
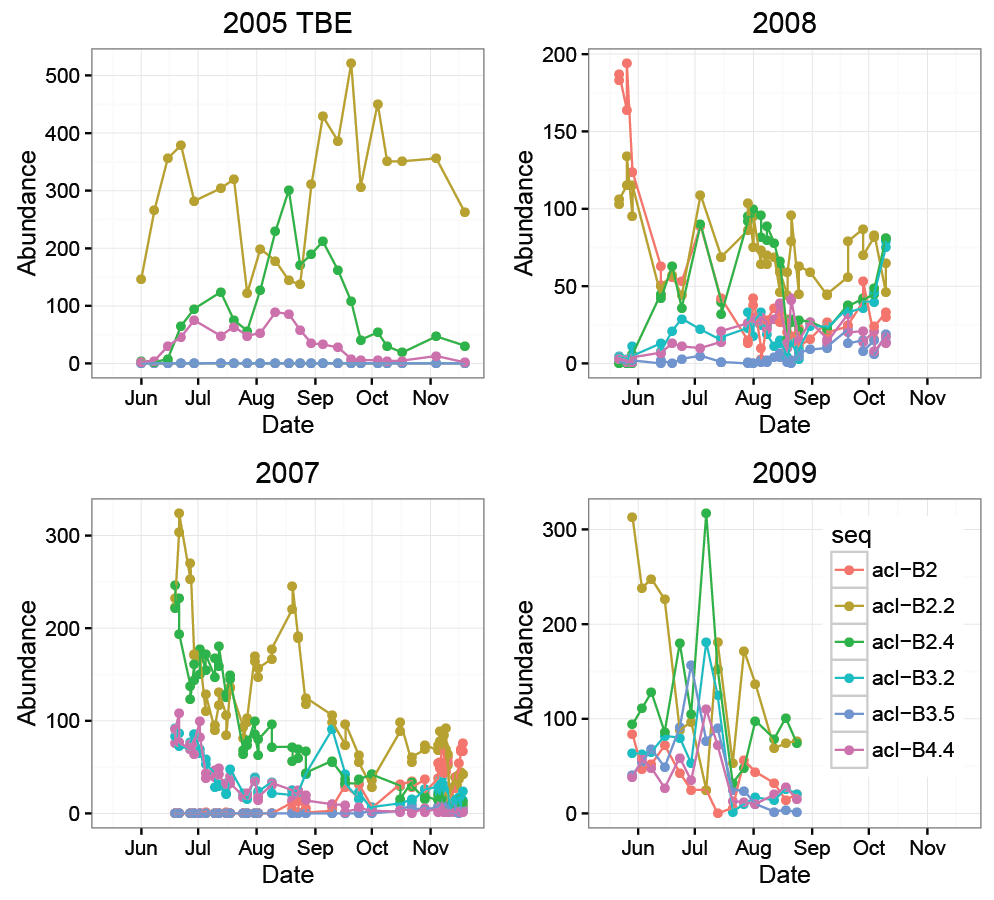


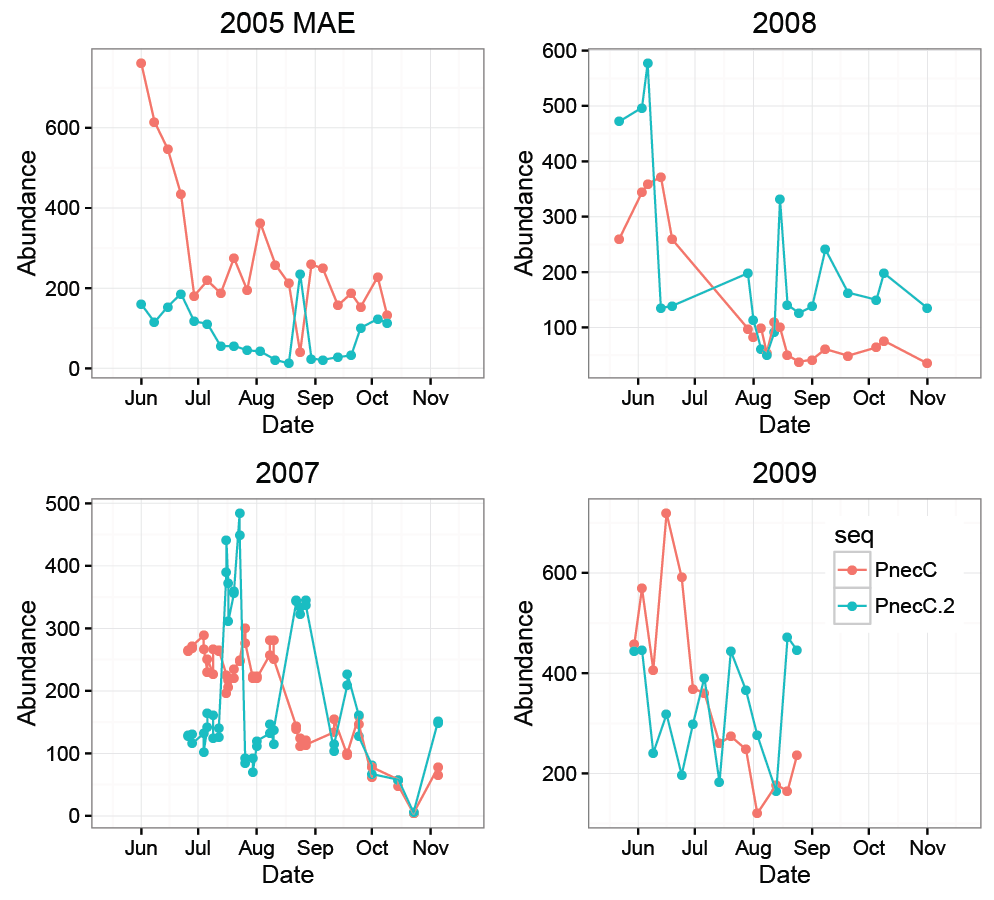


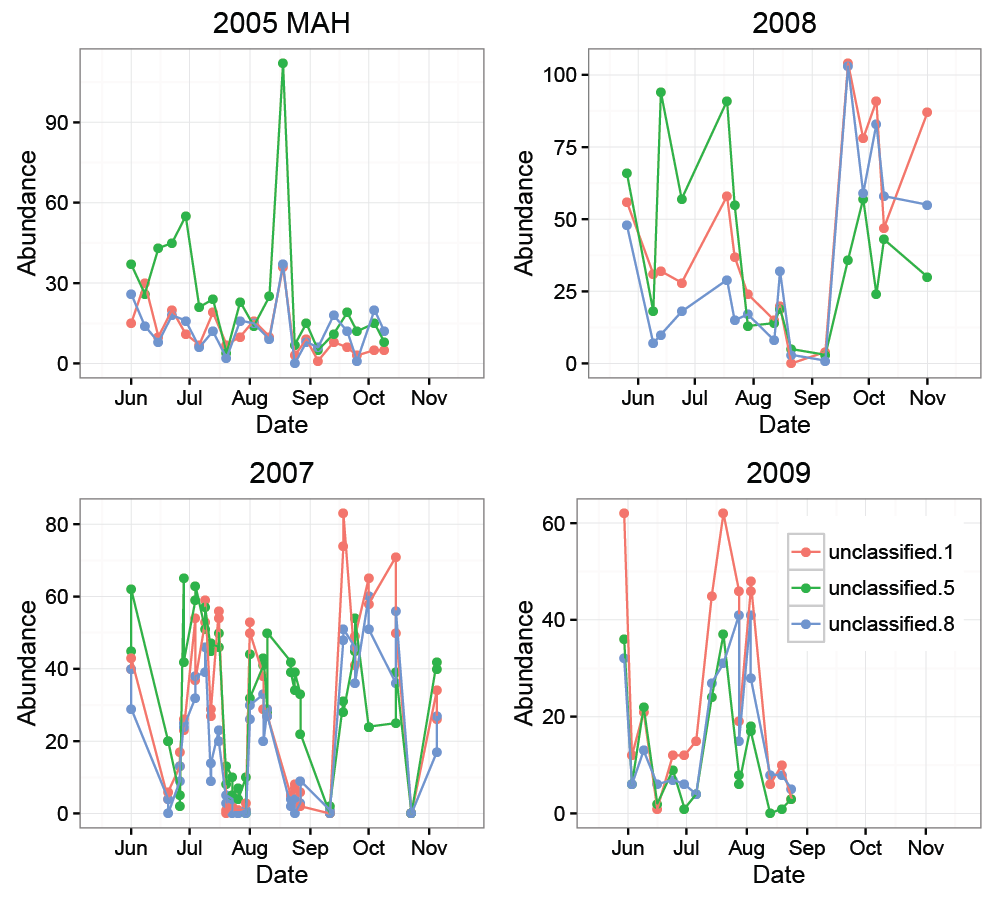
**Figure S4. Phylum composition by month by lake.** Seasonal trends, largely driven by Proteobacteria, are found in nearly all lakes and layers. Sites with multiple years of sampling are shown above. Most sites show a slight increase or no clear trend in the proportion of Proteobacteria over time, with the exception of Crystal Bog, which shows increased proportions of Proteobacteria over time. The Mary Lake hypolimnion, which has no recorded mixing events, has a stable community composition at the phylum level over seasons.



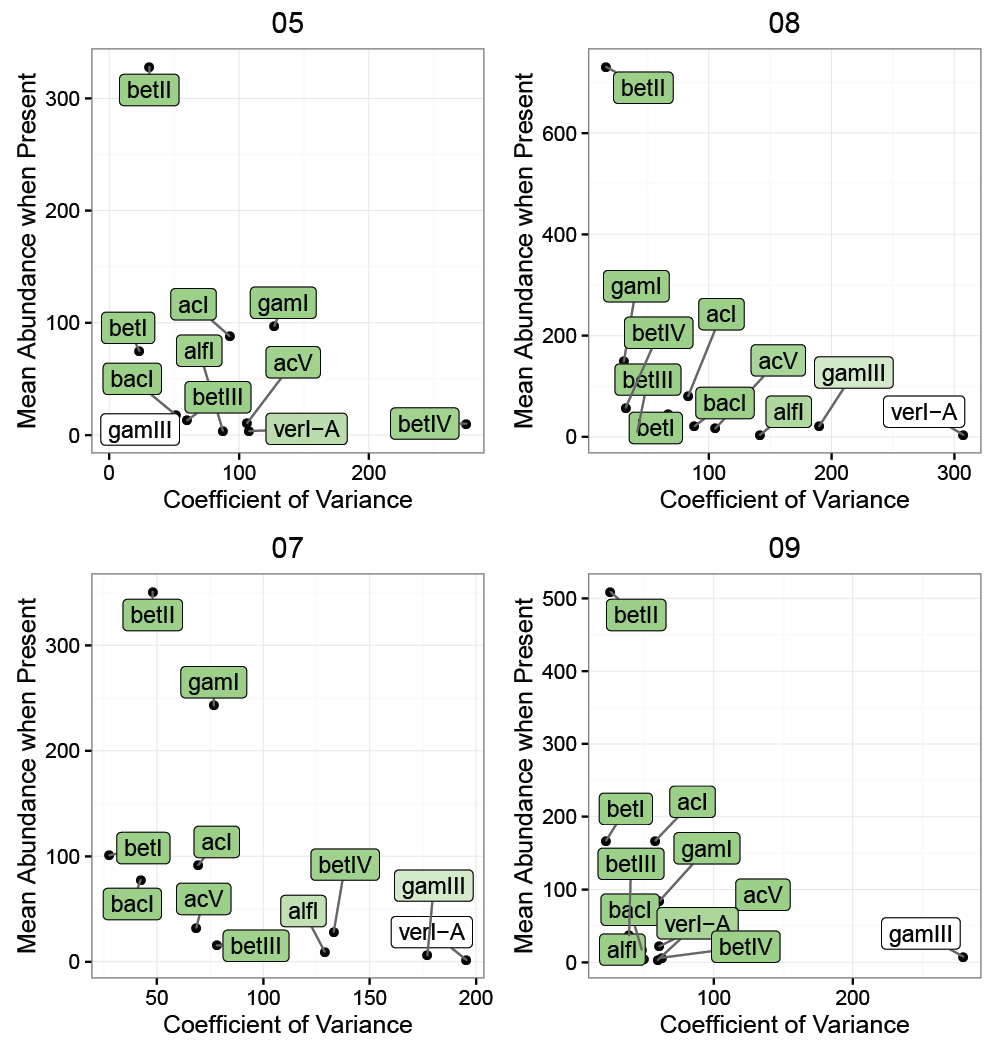
**Figure S5. Phylogenetic clustering by lake.** The mean nearest taxon distance quantifies the phylogenetic relatedness of taxa in a species compared to a null model. Negative values indicate underdispersal – that is, more phylogenetic clustering than would be expected if phylogeny was randomly distributed. This indicates that environmental filtering is a strong driver of community composition in all sites. Dimictic lakes are significantly more clustered than polymictic or meromictic lakes, as tested by a pairwise Wilcoxon test with Bonferroni adjustment.



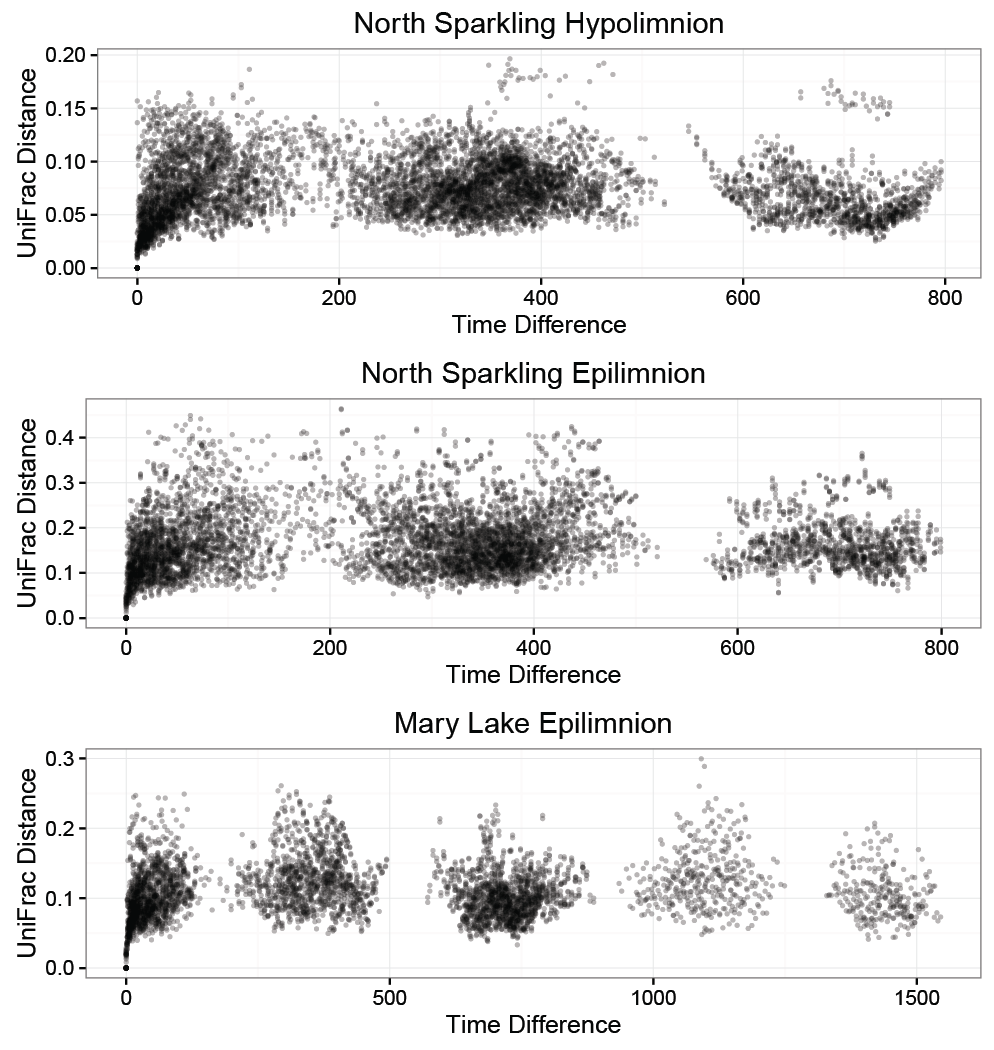


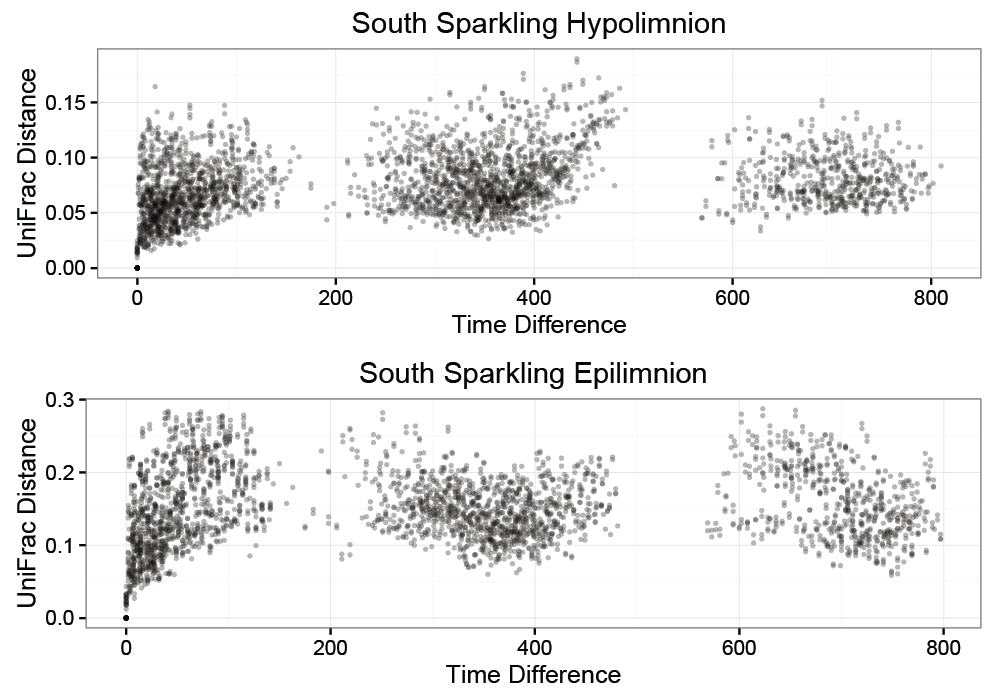


**Figure S6. Trends over time in OTUs.** Additional abundance trends of closely related OTUs by year are shown above. In Trout Bog, both acI-B and PnecC have different trends amount OTUs classified as the same tribe, and the relationships between OTUs varies from year to year. In Mary Lake, closely related OTUs of acI and betII show highly similar abundance patterns to each other, although these trends still vary from year to year. This may be due to Mary Lake’s lack of mixing disturbance and strong vertical gradients.



**Figure S7. Lineage traits by year.** Figure 5 demonstrates that lineages show consistent traits in different lakes; this plot shows that those traits are relatively consistent between years as well, as in the Trout Bog hypolimnion.





**Figure S8. Additional plots of UniFrac distance vs time between samples.** As seen in Figure 3, there is no increase in similarity between samples taken one year or multiples of one year apart. Trends appear logarithmic, with similarity between samples rapidly decreasing at first, then approaching an asymptote. This indicates that community composition does not repeat annually.