## Introduction

One of the major goals of microbial ecology is to predict bacterial community composition. The need for prediction can be demonstrated in situations such as the connection between disease and the human microbiome (1), the detrimental impacts of cyanobacterial blooms in freshwater (2), and a variety of industrial applications including chemical production (3), biofuel production (4), wastewater treatment (5), and food science (6). However, we have only a cursory knowledge of the factors that would allow us to predict 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.

Long-term studies of the bacterial communities are less common due their logistical difficulties and the need for stable funding, but results from the Microbial Observatory projects are promising. As a few examples among many, the San Pedro North Pacific - Microbial Observatory contributed to our understanding of heterogeneity of bacterial communities across space and time (7), while research at the Sapelo Island – Microbial Observatory has led the field in linking genomic data to metadata (8) and in the sampling and analysis of environmental RNA (9). In our own North Temperate Lakes – Microbial Observatory, based in Wisconsin, USA, a multi-year time series of metagenomic data was used to study sweeps in diversity at the genome level (10), adding to our knowledge of how genetic mutation influences bacterial communities. Long-term microbial ecology studies have a time-tested role in the quest to forecast bacterial communities.

In this research, we present 16S amplicon data from the North Temperate Lakes – Microbial Observatory. This dataset includes over 1,300 samples, spanning four years, eight lakes, and two depths. The lakes included in this study are all bog lakes in the boreal region of northern Wisconsin near Minocqua. Bog lakes are open regions of water surrounded by a slowly expanding mat of sphagnum moss, which leaches large amounts of dissolved organic carbon into the water in the form of humic and fulvic acids. These ecosystems contain much greater amounts of carbon than of nitrogen or phosphorus, earning the categorization of “humic” or “dystrophic.” Because the dissolved organic carbon stains the water dark brown, these lakes quickly absorb heat from sunlight and form strong vertical gradients structured by temperature and dissolved oxygen concentrations. Depending on depth, surface area, and bathymetery, bog lakes mix at different frequencies. 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).

The well-defined boundaries and environmental gradients of bog lakes make them 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. Some estimates place freshwater carbon emissions equivalent or even greater to those from the ocean. These greenhouse gases are produced and consumed by the bacterial communities. Understanding how bacterial communities change over time could improve our understanding of carbon cycling in freshwater, and therefore our global estimates of natural carbon emission and capture.

We use our Microbial Observatory 16S dataset to analyze bacterial community composition along both space and time. The communities within each lake and layer are compared, and taxa with preference for certain environmental conditions are identified. We also investigate seasonal trends, and find that broad metrics 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. The results of our analysis can be used to guide future attempts to predict bacterial community composition.

## Methods

**Sample Collection**

Water was collected from eight bog lakes during the summers of 2005, 2007, 2008 and 2009, as previously described (11). 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 the laboratory, 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 FastDNA Spin Kit for Soil (MP Biomedicals, Santa Ana, CA), with minor modifications (12) The sampling sites are located near Boulder Junction, WI, and were chosen to include lakes represent all three mixing regimes. (Table 1). Three lakes are polymictic, three are dimictic, and two are meromictic. Trout Bog and Crystal Bog are 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 platforms (buoys) on Trout Bog and Crystal Bog, allowing for more 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 | 46.047776, -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 |
| 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, WI, in the southern boreal region of North America. 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 (13). 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.

570 sequences with long homopolymer runs, ambiguous base calls, or incorrect sequence lengths were found and removed via mothur v1.34.3 (14). Thirty-three chimeras and 340 chloroplast sequences (based on pre-clustering and classification with the Greengenes 16S database, May 2013) (15) were removed. The remaining 8,913 unique sequences were clustered into 98% OTUs using mothur’s average neighbor clustering algorithm. 98% OTUs were created instead of 97% OTUs to maintain consistency with the established taxonomy for freshwater bacteria, in which a “tribe” represents a population with 98% sequence similarity (16). 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 in order 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 (16) or the Greengenes database based on the output of NCBI-BLAST (blast+ 2.2.3.1) (17). 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 (18), 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.). Significant differences in observed richness between each pair of lakes were identified using the Wilcoxon signed-rank test implement using “exactRankTests” (Hothorn and Hornik (2013). exactRankTests: Exact Distributions for Rank and Permutation Tests. R package v0.8-28). This is a non-parametric alternative to the Student’s t-test for non-normal data. Evenness was measured using Pielou’s Evenness Index. Both alpha diversity and evenness were measured at the OTU level.

Similarity between samples was compared using UniFrac distances, as implement in “phyloseq” (19) (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” (20). 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 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 code from this study can be downloaded from the R package “OTUtable” and the McMahon Lab GitHub repository “North\_Temperate\_Lakes-Microbial\_Observatory.”

## Results

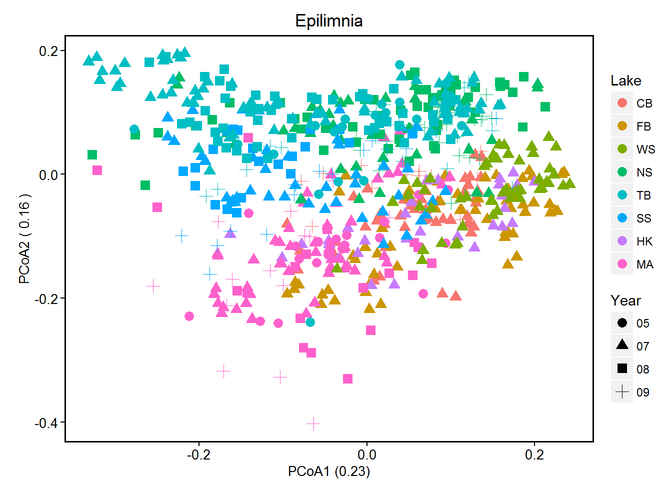
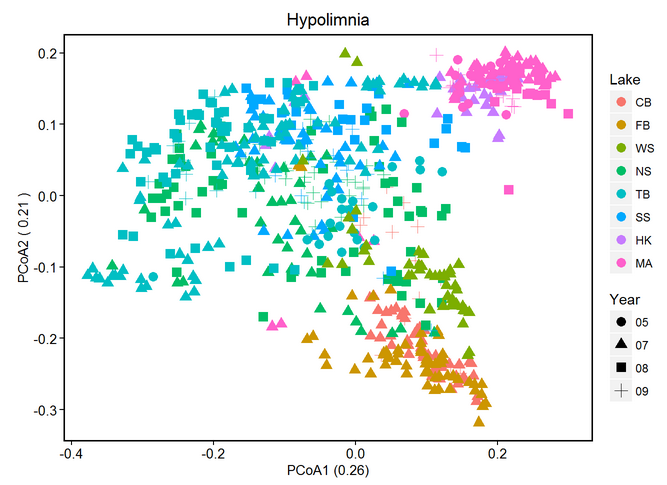
**Overview of community composition**

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 *Verrucomicrobia* were the most abundant phyla. Within these phyla, OTU abundance is highly uneven. For example, much of the abundance of *Proteobacteria* can be attributed to OTUs belonging to the well-known freshwater groups *Polynucleobacter* and *Limnohabitans,* and the freshwater clade acI contributes disproportionately to 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.

**Clusters of community composition**

When differences in community composition are quantified using weighted UniFrac distance, several trends emerge. Communities from the epilimnion and hypolimnion layers are distinct from each other in all lakes. 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. OTUs belonging to candidate phyla are more frequently observed in hypolimnia. These differences are less pronounced in lakes that mix frequently, and likely reflect the bias in freshwater microbial ecology towards the surface waters of lakes. Within layers, mixing regime is the next driver of community composition. 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 (West Sparkling Bog and Forestry Bog) in 2007, but appears more similar to dimictic hypolimnia in 2009. High resolution buoy data for Crystal Bog is available through the North Temperate Lakes Long Term Ecological Research Project, and this data shows that Crystal Bog mixed multiple times in 2007, but only twice in 2009. (Check 07 and 08 dates of NS points in the polymictic cluster)

The stark contrast between years in Crystal Bog is mirrored to a lesser extent in the other lakes in this study. Each year of sampling from the same lake is unique and distinct, particularly in hypolimnia. 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 values from MANOVA). This suggests that environmental filtering is occurring based on parameters unique to each site.

**Indicator taxa**

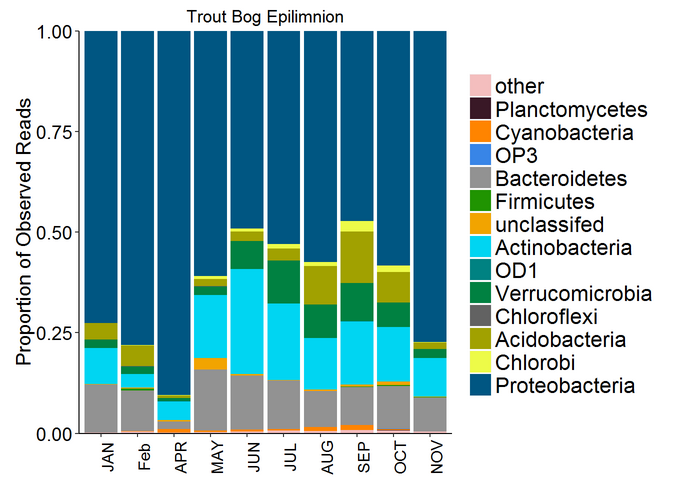
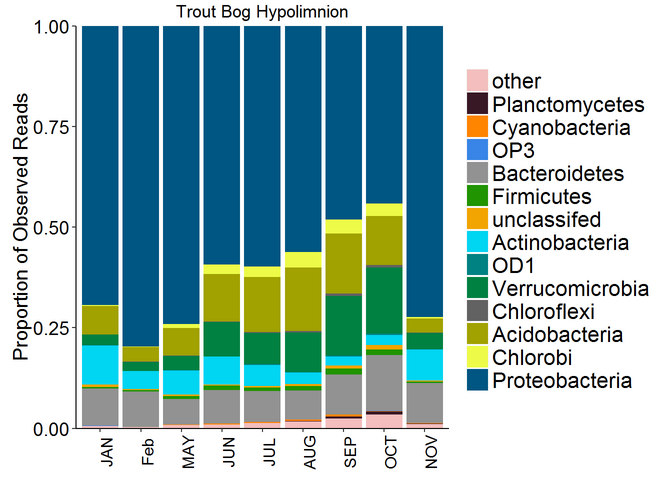
To further investigate potential drivers of the differences in community composition between mixing regime, we performed indicator analysis (Table 2). This technique identifies taxa that are found significantly more often in one group of samples than another; in this case, the groups were defined by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

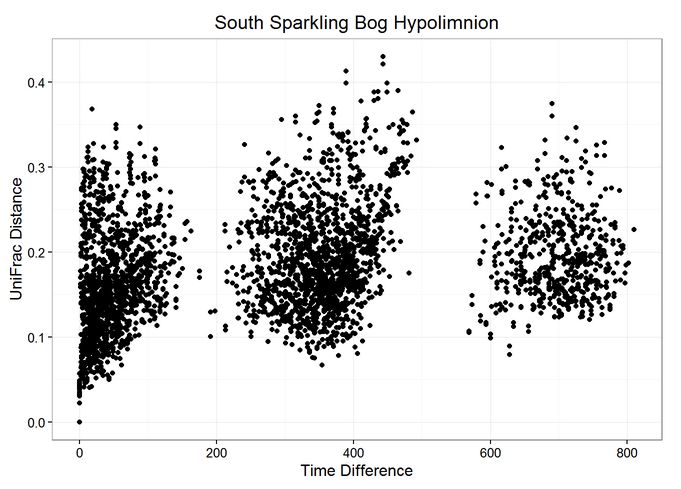
**Table 2. Indicator taxa of layers and mixing regimes.** Indicator analysis was used to identify taxa with a preference for sites defined by layer and mixing regime. OTUs were grouped into higher taxonomic levels, and all levels were used for this analysis at once. The lowest classification of each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

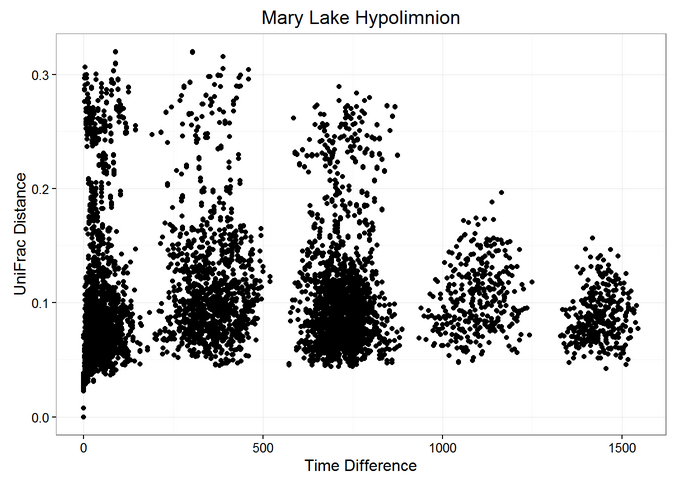
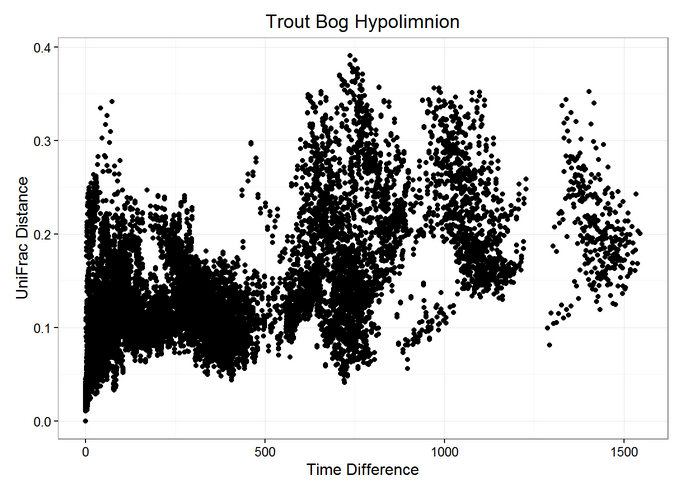
|  |
| --- |
| **Polymictic epilimnia** |
| 1. Actinobacteria;Actinobacteria; Actinomycetales;acI;acI-B;acI-B3 2. Verrucomicrobia;Spartobacteria; Chthoniobacterales;verI-A 3. Proteobacteria;Betaproteobacteria;Burkholderiales;betI;betI-A;Lhab-A4 4. Proteobacteria;Betaproteobacteria;Methylophilales 5. Actinobacteria;Actinobacteria;Actinomycetales;acI;acI-B;acI-B3 6. Proteobacteria;Betaproteobacteria;Burkholderiales;betI;betI-A 7. Verrucomicrobia;Spartobacteria 8. Actinobacteria;Actinobacteria;ActinomycetalesMycobacteriaceae;Mycobacterium 9. Actinobacteria;Actinobacteria;Actinomycetales;acI 10. Proteobacteria;Betaproteobacteria;Burkholderiales;betII;Pnec;PnecC |
| **Dimictic epilimnia** |
| 1. Proteobacteria;Betaproteobacteria;Methylophilales 2. Actinobacteria;Acidimicrobiia;Acidimicrobiales;acV 3. Actinobacteria;Actinobacteria;Actinomycetales;acI;acI-B;acI-B2 4. Bacteroidetes;Saprospirae;Saprospirales;Chitinophagaceae 5. Bacteroidetes;Sphingobacteriia;Sphingobacteriales;bacVI;bacVI-A;Muci 6. Proteobacteria;Betaproteobacteria;Burkholderiales;betI;betI-B;Rhodo 7. Acidobacteria 8. Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae 9. Chlorobi;Chlorobia 10. Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae |
| **Meromictic epilimnia** |
| 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 9. Proteobacteria;Alphaproteobacteria;Rickettsiales 10. Proteobacteria;Betaproteobacteria;Burkholderiales;betI;betI-A;Lhab-A1 |
| **Polymictic hypolimnia** |
| 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 9. Actinobacteria;Acidimicrobiia;Acidimicrobiales 10. Proteobacteria;Alphaproteobacteria |
| **Dimictic hypolimnia** |
| 1. Acidobacteria;Holophagae;Holophagales;Holophagaceae 2. Acidobacteria 3. Proteobacteria;Betaproteobacteria;Burkholderiales;betII;Pnec;PnecC 4. Proteobacteria;Betaproteobacteria;Methylophilales 5. Verrucomicrobia;Pedosphaerae;Pedosphaerales;Ellin515 6. Proteobacteria;Betaproteobacteria 7. Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae;Desulfobulbus 8. Proteobacteria;Betaproteobacteria;Burkholderiales;betI;betI-B;Rhodo 9. Proteobacteria;Betaproteobacteria;Methylophilales;Methylophilaceae 10. Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobulbaceae |
| **Meromictic hypolimnia** |
| 1. Proteobacteria;Deltaproteobacteria;Syntrophobacterales 2. Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophaceae 3. OP3;BD4-9 4. OP3 5. Verrucomicrobia;Verruco-5;WCHB1-41 6. Proteobacteria;Deltaproteobacteria;Desulfobacterales;Desulfobacteraceae 7. Planctomycetes 8. Archaea;Parvarchaeota 9. Verrucomicrobia;Verruco-5 10. OP8;OP8\_1 |

**Seasonal trends**

Despite distinct community compositions each year, there are high level seasonal trends. Both biodiversity and evenness increase over time while stratification is in place (supplemental). Sharp decreases in both of these community properties are observed during the well-sampled Trout Bog 2007 fall mixing event and the North Sparkling Bog artificial mixing event. At the phylum level, the abundance of *Proteobacteria* is particularly high in the few winter samples in this dataset and during mixing events. After spring stratification, the abundance of *Proteobacteria* decreases 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.

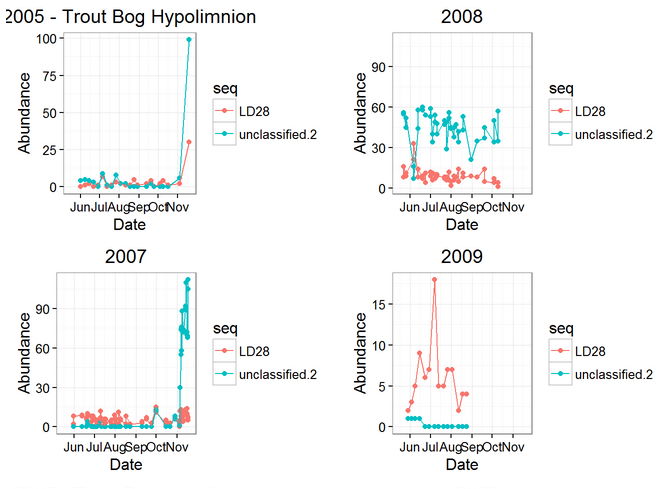
**Figure 2. Phylum-level seasonal trends.** The proportion of Proteobacteria is highest in winter, early spring, and fall, but decreases during the stratified summer months. Most other phyla show the opposite trend. Because this is based on relative abundance data, this trend likely reflects the observed increase in biodiversity with time, and adds that members of Proteobacteria are even more dominant when bog lakes are not stratified.

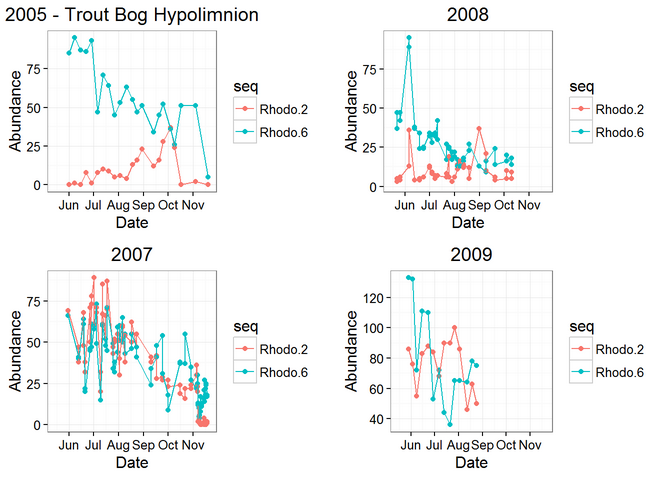
However, seasonal trends break down at lower taxonomic levels. In other aquatic ecosystems, OTUs show repeatable trends in abundance each year, resulting in high similarity between samples collect during the same season but different years. This is frequently depicted by plotting a beta diversity metric such as UniFrac distance between samples versus the time between sample collection. This generally produces a sine wave-like pattern with decreasing amplitude. When we perform the same analysis on this dataset, 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 1 year and 2 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.



**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.

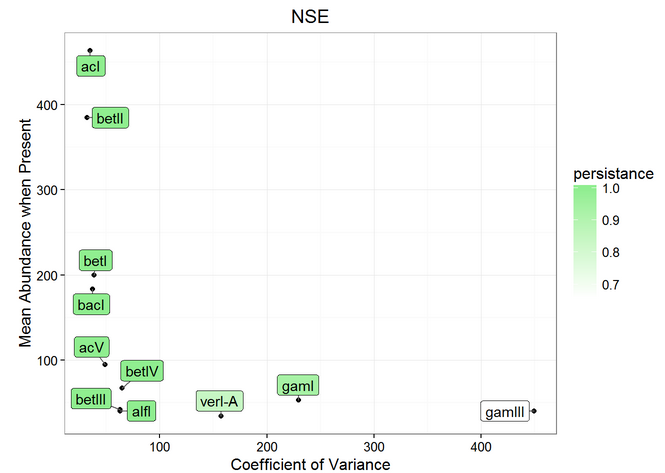
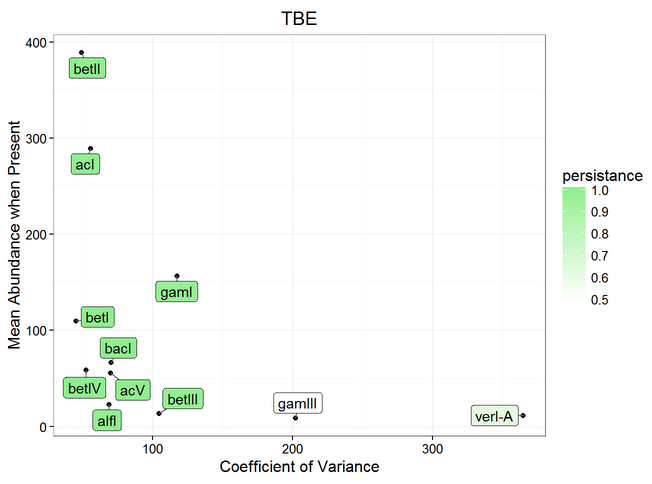
**Trends in specific OTUs do not repeat annually**

 As suggested by the lack of similarity in community composition between samples taken one year apart, trends in OTU abundance over time do not repeat annually. 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. The reasons for these observations are unclear; with only four years of data, differences between years cannot be linked to climatic factors, events in the surrounding landscape, or changes in nutrient levels. It is also possible that biotic drivers and stochastic community assembly are at play in these ecosystems.

**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 for clades classified using the freshwater taxonomy. Using only these well-defined freshwater groups allowed better taxonomic resolution. This analysis showed that low persistence is associated with high variability, and that low variability is associated with high abundance. We rarely observe “bloomers,” situations where a clade has both high abundance and low persistence. Most freshwater clades are highly persistent at low abundances with low variability. Clade gamIII of the Gammaproteobacteria is an exception, with low persistence, low abundance, and high variability. Clades gamI and verI-A occasionally also exhibit this profile. Clades betII and acI are highly abundant and persistent with low variability, consistent with their profile as ubiquitous freshwater generalists. Knowledge of the general manner in which these clades behave can begin to shed light on their lifestyles and the reasons for their observed abundance trends.



**Figure 5. Traits of freshwater clades.** These well-defined freshwater clades show similar persistence, 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 variability, freshwater clades themselves have predictable traits. We see that persistence, abundance, and coefficient of variance of freshwater groups are relatively consistent between years and lakes. If this is the case, then why is there such strong differences between years?

Based on the large differences in community composition between epilimnia and hypolimnia, and between lakes, environmental filtering likely has a strong effect in bog lakes. This is supported by increased phylogenetic clustering in all samples compared to what would be expected if phylogeny was random (supplemental). This is supported by the identity of indicator taxa for mixing regimes. 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 generalist acI are indicators exclusively for epilimnia and polymictic hypolimnia, where they can likely take full advantage of sunlight to power actinorhodopsins. But while the presence of taxa endemic to specific sites is consistent from year to year, their abundance trends are highly variable.

One possibility is that there are abiotic drivers not measured that vary strongly from year to year. These bog lakes, being small compared to other studied freshwater systems such as Lake Mendota in Madison, WI, are more at the mercy of their surrounding landscape. Changes in the terrestrial ecosystem surround the aquatic systems studied may have profound impacts on bacterial community composition. Another potential driver of annual variation is biotic interactions. Bacterial communities have been shown to take part in complex networks, and forces such as competition and cooperation may be structuring community composition in addition to environmental filtering. Finally, stochastic community assembly may occur after the disturbances of mixing events, the winter ice-on period appears to be a bottleneck reducing the alpha diversity of the bacterial community. 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.

If it is true that repeatable annual trends do not repeat in bog lakes, why has this not been seen in other aquatic ecosystems? Previous studies in these same lakes found seasonal trends, but were using higher level community traits such as richness or rate of change rather than direct community composition. There is a large body of work on predictable seasonal trends in oceans, rivers, and freshwater lakes. One major difference 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. Bacterial communities that are constantly moving through an ecosystem may experience stronger environmental filtering, reducing variability and the importance of biotic interactions in community assembly. 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. 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. Additionally, the Mendota study notes that late summer samples are more difficult to predict than fall, winter, or early spring samples. Since the bog lake dataset is dominated by stratified summer samples, this would make annual prediction far more difficult. (Include Martin Hahn’s polynucleobacter study here)

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 drastically 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’s 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 sequencing. 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 environmental condition. Seasonal 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 lower taxonomic levels. Each year in each lake harbors a unique bacterial community, within the confines of its environmental conditions. 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 variation are consistent in freshwater taxa despite large annual differences. Interannual variability could be due to either environmental factors that operate on long time scales, or due to 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.

To-Do

* Make figures more professional
* Add citations
* Write an abstract
* Make a supplemental document

## References

1. **Cho I**, **Blaser MJ**. 2012. The human microbiome : at the interface of health and disease **13**.

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. No Title **686**.

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.

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. International Journal of Food Microbiology Food fermentations : Microorganisms with technological bene fi cial 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 1–13.

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

12. **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.

13. **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.

14. **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.

15. **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.

16. **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.

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

18. **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.

19. **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.

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