Here, we present the North Temperate Lakes - Microbial Observatory dataset, a multi-year, multi-lake time series. We found that multiple years of sampling were necessary to describe the community of bog lake ecosystems, and that bacterial community richness and membership was structured by layer, mixing regime, and lake. We identified specific bacterial taxa present throughout the dataset, as well as taxa endemic to certain depths or mixing regimes. Mixing events, marked by reduced biodiversity, also selected for specific taxa. Despite these trends in membership, high levels of variability were detected in this dataset; each year in each lake harbors a unique bacterial community, and that community is often dispersed over a single year. Our results emphasize the importance of long term time series in microbial ecology in order to assess bacterial community membership and variability.

The bog lakes used 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. Our research builds upon these studies and adds information about the taxonomic identities of bacterial groups. For example, persistent and unique bacterial groups were detected in the bog lakes using ARISA; using 16S amplicon sequencing, we have determined that the ubiquitous freshwater bacteria LD28, acI-B2, PnecC, and bacI-A1 are core members of the bog lake community. Differences in richness and community membership were detected in Crystal Bog, Trout Bog, and Mary Lake, three sites representative of the three mixing regime categories of polymictic, dimictic, and meromictic. Our data supports these results and suggest that these trends are indeed linked with mixing regime, as we included multiple lakes of each type in this study. Synchrony in the rate and direction of change has been extensively studied in these lakes; because the structure of our dataset includes variable intervals between sampling and different numbers of samples each year, we cannot support or disprove the hypotheses about concordance in bog lakes. We can confirm that epilimnia tended to be more variable than hypolimnia. We also supported previous research on the impacts of lake mixing on bog lake communities. Mixing did appear to be disruptive to both epilimnion and hypolimnion communities, selecting for only a few taxa that can thrive during this disturbance, but quickly recovering diversity. Comparing richness between lakes of different mixing regimes did not support the intermediate disturbance hypothesis, though. Instead, the least mixed lakes had the most diverse communities. As many variables are co-dependent with mixing regime (such as depth, volume of integrated water column, dissolved carbon concentrations and total nitrogen concentration), it is not clear what is driving this trend.

Despite including multiple years of sampling, we were not able to detect seasonal trends in bog lakes. While seasonality in marine and river systems has been well-established by our colleagues, previous research on seasonality in freshwater lakes is mixed. Nelson identified distinct, repeatable community types throughout the year in alpine lakes, but noted that stratified summer communities were distinct each year. Seasonal trends were detected in a similar time series from Lake Mendota, but summer samples were more variable then those collected in other seasons. In the bog lakes included in this study, community properties such as richness and rate of change were consistent each year, and the phytoplankton community was observed to drive seasonal trends in the bacterial community. However, a second year of sampling for seasonal trends in Crystal Bog was not able to repeat these findings. Another study comparing successional trends in three of the bog lakes describes “dramatic jumps” in community composition during the summer months, while spring, winter, and fall had more stable community composition. Because the majority of our dataset was collected during the summer stratified period, increased summer variability may explain why we see a different community each year and a lack of seasonal trends. However, we cannot disprove the presence of seasonality in temperate freshwater lakes.

One of the biggest benefits of 16S data over ARISA is the ability to classify sequences. 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 biogeochemical differences driven by mixing regime; for example, dimictic and meromictic hypolimnia, which are consistently anaerobic, harbor putative sulfur reducers not present in polymictic hypolimnia, which are more frequently oxygenated. Members of clade acI partition by mixing regime in epilimnia, though the functional traits driving this filtering effect as of yet unknown. Interestingly, the hypolimnion of meromictic Mary Lake contains several taxa classified into the candidate phyla radiation and a larger proportion of completely unclassified reads. This is consistent with other 16S and metagenomics studies of meromictic lakes, and suggests that the highly reduced conditions in these systems support taxa not detected in more commonly sample environments.

Perhaps the biggest implication of this study is the importance of long-term time series. A similar dataset spanning only a single year would not have captured the full extent of variability observed, and therefore would not have capture as many of the taxa belonging to the bog lake community. 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. Biotic interactions between bacterial taxa may also contribute to interannual variability. Understanding the factors that contribute to a lack of repeatable seasonal trends in bog lakes will lead to improved predictive modelling in freshwater systems, allowing forecasting of bloom events and guiding better management strategies. Additionally, these highly variable systems may be ideal for addressing some of the core questions in microbial ecology, such as how community assembly occurs, how interactions between taxa shape community composition, and how resource partitioning drives the lifestyles of bacterial taxa.

To answer these questions and more, 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.