From microbes to ecosystems: time series provide insight into how microbial metabolisms scale to ecosystem functions

By

Alexandra Linz

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Microbiology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: July 31, 2018

The dissertation is approved by the following members of the Final Exam Committee:

Katherine McMahon, Professor, Bacteriology, Civil & Environmental Engineering
Timothy Donohue, Professor, Bacteriology
Anthony Ives, Professor, Zoology
Garret Suen, Associate Professor, Bacteriology
Stefan Bertilsson, Professor, Ecology and Genetics

Abstract

The study of freshwater microbial ecology has the potential to improve our knowledge both of biogeochemical cycling in freshwater and of the ecology of microbial communities. However, this area of research also presents several challenges. The first is that most freshwater microbes cannot yet be cultivated in the lab. While the number of freshwater isolates is steadily increasing, the high levels of diversity observed in freshwater microbial communities make describing biogeochemical cycling one species at a time a herculean task. The second is the difficulties of experimentation in lakes - designated control or replicate lakes often make poor comparisons, mesocosm experiments are subject to bottle effects, and there are high levels of variability within these communities. In this thesis, I use time series of DNA and RNA sequencing data to investigate uncultured freshwater microbes, to view the microbial communities as whole, dynamic entities, and to provide statistical power to a system that is difficult to control and replicate.

Using a multi-year time series of 16S rRNA amplicon data, I identify characteristic community members of bog lakes and assess their variability. By analyzing a multi-year metagenomic time series and genomes generated from that data, I compare microbial metabolisms between two lakes to infer differences in biogeochemical cycling. Finally, I use a two day time series of metatranscriptomics to learn how gene expression changes in day versus night and to learn how closely related taxa differentiate their niches and co-exist. This work adds to our knowledge of freshwater microbial ecology and can be used as a resource by other researchers.

Table of Contents

Acknowledgements

- Thanks to my advisor, Trina McMahon, who has always encouraged me to do my best work and leave no stone unturned. She has inspired me to tackle challenges outside of my area of expertise, and I know I have become a better scientist and a more confident person because of her.

- Thanks to my committee members, Tim Donohue, Garret Suen, Tony Ives, and Stefan Bertilsson, who have gone above and beyond what is expected to advise me on science and my career. I've always appreciated that committee meetings have had a friendly and relaxed atmosphere.

- Thanks to all my labmates for their support over the years, whether it was in the form of help in the field, teaching me a new technique, or just listening when it seemed like the scientific method had stopped working. Especially Sarah Stevens, who I will probably still pester with coding questions after we graduate.

- Thanks to all the undergraduates I have worked with: Ben Kranner, Sam Schmitz, Maggie Sobolewski, and Kaela Amundson. I think I learned more from them than they did from me, and I know they will all go on to do great things.

- Thanks to everyone at UW Trout Lake Station for going out of their way to help me with field work - especially Tim Meinke, who single-handedly saved the work in Chapters 5 and 6 by giving me his last specialty data logger battery when mine died in the middle of sampling.

- Thanks to the LTER community for giving me a crash course in limnology and ecology, for volunteering to collect samples with me in the middle of the night, and for lending me some terrifyingly expensive equipment.

- Thanks to my family - my parents, my sister, my grandparents, and my in-laws - for providing emotional support, keeping me grounded, and making sure I'm getting enough to eat. They may not be 100% sure what I do, but they're probably the most dedicated research audience I have.

- Most of all, thanks to my wonderful husband, Dan. You encourage me when times are tough and celebrate with me when times are good. You've been there every step of the way, whether it's proofreading my applications, making me hot chocolate after I fall in the lake, or bringing me dinner during late nights in the lab. I'm so glad you're my life partner!

I couldn't have done it without you!

List of Figures and Tables

Chapter 1: Introduction

Microbes in freshwater ecology

Microbes form the foundations of ecosystems, and the importance of the functions they perform far outweighs their cell sizes. They are responsible for fixing carbon and nitrogen from the atmosphere, recycling nutrients from higher trophic levels through decomposition, transforming inorganic nutrients such as nitrogen, phosphorus, and sulfur, and adding energy to ecosystems via photosynthesis. Without microbes, ecosystems could not support multicellular life or many of the ecosystem services on which humans depend.

Freshwater ecosystems are both vital to human health and susceptible to issues caused by changes in their resident microbial communities. Beach closings and contaminated drinking water supplies are high profile examples of how microbes directly impact freshwater ecosystem services, but they can also have indirect effects, such as fish kills due to microbially-induced hypoxia (1) and the methylation of mercury to its more toxic form (2). Because microbes are at the base of the freshwater food chain, seemingly minor changes in the microbial community can ripple throughout the entire ecosystem.

Freshwater receives nutrients from surrounding terrestrial ecosystems, making lakes, ponds, and rivers “hot spots” of biogeochemistry at the landscape level (3). Much

of this chemical processing is performed by microbes, which transform compounds largely inaccessible to higher trophic levels into consumable biomass. The connectivity of freshwater within landscapes makes nutrient cycling in these ecosystems an important area of research (4). However, microbes are often considered to be a single, stable entity for the purposes of ecosystem nutrient budgets, an assumption that has been proven false in many systems and may lead to incorrect conclusions (5–7). Therefore, the study of microbial community dynamics in freshwater has the potential to drastically improve our understanding of freshwater carbon and nutrient cycling.

A word on terminology

This thesis is located at the intersection of microbiology and limnology. These two fields have approached microbes in freshwater differently for some time, and often used conflicting terminology. In this thesis, I generally prefer to use the microbiology terms, as these are broadly applicable to other ecosystems and allow comparisons with the body of literature on microbiomes. However, knowledge of both terms is necessary to synthesize previous research from both fields.

An important distinction in limnology is the classification of organisms into *nekton*, *plankton*, and *benthos*. Nekton are organisms that can move more quickly than water currents; in lakes, this typically refers to fish. Plankton include organisms such as zooplankton, algae, and larvae that cannot move as quickly as the currents and are at the mercy of physical limnology. Benthos are organisms that inhabit the lakebed and sediments. Water column microbes are categorized as plankton in lakes. Plankton can

be further classified by size, with microbes typically considered *pico-* to *microplankton*, or by trophic level. Photosynthetic plankton are termed *phytoplankton,* a group that includes tiny plants, eukaryotic algae, and phototrophic prokaryotes such as *Cyanobacteria*. "Algae" is occasionally used as a synonym for phytoplankton, a name contrary to microbiology, where algae typically refers to eukaryotes only. *Zooplankton* are small animals or protists that consume phytoplankton or detritus. Classically, these were the only trophic distinctions, but more recently, *bacterioplankton* has been used to describe heterotrophic bacteria, while *mycoplankton* refers to heterotrophic fungi.

In microbiology, metabolism types define freshwater microbial groups. *Autotrophic* organisms gain carbon through carbon fixation, while *heterotrophic* organisms meet their carbon requirements by ingesting organic compounds. *Phototrophic* organisms use light as an energy source, and *chemotrophic* organisms use chemical energy. Combinations of these classifications are termed *photoautotrophs, photoheterotrophs, chemoautotrophs,* and *chemoheterotrophs*. Photoautotrophic organisms are roughly equivalent to phytoplankton, while chemoheterotrophs could also be called bacterioplankton. Autotrophs and heterotrophs could be said in limnology terms to be performing *primary production* (adding carbon to the ecosystem by fixing it from the atmosphere) and *respiration* (producing carbon dioxide that is released from the ecosystem to the atmosphere), respectively. Note that respiration has a different meaning in microbiology, where it is used to refer to the generation of energy via an electron transport chain. Conveniently, the most of the metabolic pathways leading to

microbial respiration produce carbon dioxide as a byproduct, so this terminology conflict does not often become an issue. It is unclear where photoheterotrophs and chemoautotrophs fit into the limnology classification scheme. Therefore, I define freshwater microbes by their metabolism types in this thesis.

Similarly, "nutrients" can be a conflicting term in ecology and microbiology. In microbiology, nutrients refers to any chemical requirement of a microbe, whether it is carbon, nitrogen, a terminal electron acceptor, or a vitamin. For example, a rich media used for culturing may be called a nutrient broth. In ecology, a nutrient is typically an element such as nitrogen, sulfur, or phosphorus that is transformed within an ecosystem - these would likely be called essential elements by a microbiologist. These transformations are collectively referred to as "nutrient cycling," a term I use here when referring to ecosystems.

Table 1.1 Synonymic terms from limnology and microbiology. Some key concepts in microbiology and limnology have different names. Terms that mean the same thing in both fields are listed here, along with their definitions.

Limnology Term	Microbiology Term	Definitions
plankton	microbial community	organisms that are moved by water currents; organisms too small to see by eye
phytoplankton	photoautotrophic community	microbes that perform photosynthesis and fix carbon
bacterioplankton	heterotrophic community	microbes that rely on organic carbon sources
primary production	photoautotrophy	the process of using photosynthesis coupled with carbon fixation

Table 1.2. Conflicting terms from limnology and microbiology. Some terms are used in both microbiology and limnology, but refer to different concepts. Conflicting terms and their definitions in both fields are listed here.

Term	Limnology Definition	Microbiology Definition
algae	all photoautotrophs	eukaryotic photoautotrophs
respiration	releasing carbon dioxide from the consumption of organic carbon	mobilizing energy via an electron transport chain; often, but not always, coupled with organic carbon degradation via glycolysis and the tricarboxylic acid cycle, which releases carbon dioxide
nutrients	elements that are transformed within an ecosystem	compounds that are required for growth

Lakes are not tiny oceans

When it comes to global nutrient cycling, marine ecosystems get the bulk of the attention - after all, their surface area and volume far exceeds that of any other type of ecosystem on the planet. Freshwater ecosystems are often considered part of the terrestrial ecosystem or delivery systems of nutrients from land to sea. However, freshwater itself has been shown to be an important location of nutrient processing. Approximately half of the carbon received from the terrestrial landscape is emitted as carbon dioxide (0.2 Pg C/year) or stored (0.8 Pg C/year) by freshwater ecosystems (8). While these values are still lower than emission (7 Pg C/year) or storage (2.2 Pg C/year) rates for oceans (9, 10), the amount of carbon cycling in freshwater is still greater than

would be expected given the much smaller size of freshwater ecosystems compared to marine ecosystems. We observe a similar trend with the nitrogen cycle. For example, 20% of global denitrification is estimated to occur in freshwater, roughly equivalent to the amount of denitrification taking place in soils (22%) and about a third of the amount occuring in oceans (58%) (11).

Within freshwater, emissions of carbon dioxide and methane correlate negatively with lake surface area (12). One reason for this trend is that smaller lakes have a higher ratio of terrestrial carbon received to volume. Another is that lake stratification and the diffusion of oxygen into sediments result in oxic-anoxic interfaces, which tend to host high rates of microbial growth and metabolism (13). These factors may explain why rates of nutrient cycling in freshwater seem disproportionately high compared to their size.

Microbes are responsible for much of the nutrient cycling in both marine and freshwater systems, but the types of microbes found in these ecosystems are distinct. Marine and freshwater microbes are not closely related, and crossover colonization is rare despite frequent opportunities (14). When these transition events do occur, major changes in central carbon metabolism seem to be required for survival (15). Microbial community structuring by viruses, a hallmark of marine ecosystems, appears to function differently in freshwater (16). Given these contrasting characteristics, conclusions about microbial nutrient cycling in oceans cannot be generalized to freshwater lakes.

Microbes in the food chain

Although aquatic microbes were once considered to be exclusively decomposers or phytoplankton, their role and relative importance in the food chain has since been expanded (17). Dissolved organic carbon (DOC) is produced at every trophic level, but this carbon is often not in a form available to be consumed by secondary or tertiary trophic levels. Microbes are responsible for degrading this DOC, producing biomass, and subsequently being consumed. This process of maintaining lost DOC within the food web is known as the "microbial loop" (18). Aquatic microbes perform large amounts of respiration during the microbial loop. In some systems, microbial respiration is thought to exceed primary production, resulting in the release of excess of carbon dioxide to the atmosphere (19). The balance of respiration to primary production varies with nutrient concentrations; oligotrophic systems favor heterotrophic microbes, while eutrophic systems favor phototrophic microbes (20).

The DOC in lakes that is transferred through the microbial loop to higher trophic levels comes in many forms. Traditionally, carbon in lakes is classified as *allochthonous,* which means carbon received the surrounding terrestrial landscape, or *autochthonous,* which means carbon produced in the lake. Generally, allochthonous carbon is more complex, often including cellulose, lignin, and humic substances, while autochthonous carbon is derived from photoautotrophs and includes sugars and low molecular weight acids. However, there are compounds that contradict this trend, such as the high complexity biopolymers produced by photoautotrophs. Microbes are known to degrade

both allochthonous and autochthonous DOC, although research suggests a preference for autochthonous carbon (21). The ratio of allochthonous to autochthonous carbon degradation depends on the trophic status of the lake and the productivity of its phototrophs (22). However, carbon transformation within the microbial community muddles these categories. Allochthonous DOC can be converted to compounds similar to those found in autochthonous DOC. Stable isotope analysis can be used to distinguish carbon origins, but it is not clear if microbes still prefer autochthonous over allochthonous carbon if they are in equally labile forms. To avoid this issue, I refer to carbon compounds as high complexity or low complexity in terms of ecological niches.

Particulate organic matter (POM), in contrast to DOC, is available to be consumed by either macroscopic scavengers or by microbes. POM can be either allochthonous or autochthonous and typically comprises less of the organic carbon pool than DOC, but contains labile compounds such as protein and polysaccharides (23). It is quickly colonized by microbes, and freshwater particle-associated microbial communities have different taxa than free-living microbial communities (24). Free-living and particle-associated microbes can be distinguished via serial filtration. Particles can be captured with a 5 micron pore size, while a 0.22 micron pore size is used to capture free-living microbes (25). Processing lake water through 0.22 micron filters without prefiltration, as was done in this thesis, is assumed to include both free-living and particle-associated microbes.

Inorganic compounds, while not technically part of the freshwater food web, provide energy to chemotrophs that is utilized by other trophic levels. Common terminal electron acceptors in freshwater include sulfate, nitrate, and iron (26). Methanogenesis is another potential energy-generating metabolism, and it is often inversely correlated with sulfate reduction (27). Even organic carbon can sometimes be used as to produce chemical energy; humic substances, notoriously complex and recalcitrant to degradation, can instead act as electron acceptors (28). Microbial conversions of inorganic compounds are often just as crucial to freshwater nutrient cycling as degradation of organic compounds.

Lifestyles of freshwater bacteria

It is sometimes quoted in microbial ecology that "everything is everywhere, but the environment selects" (29). Whether every microbe is truly dispersed to every ecosystem or not, we do know that freshwater receives a steady influx of immigrant microbes from the surrounding landscape and the atmosphere (30). Yet we still see that freshwater microbial communities are distinct from other environments, and there are many taxa that seem to be found only in freshwater (31). As the bacterial taxonomy is well-defined (32) (and the taxonomy of freshwater algae and fungi is an understudied area), I will focus on the bacteria typically found in freshwater lakes. Because bacterial species cannot be defined until they are in pure culture, the freshwater bacterial taxonomy is instead based on sequence similarity in the 16S rRNA gene and is organized hierarchically into monophyletic lineages, clades, and tribes. Each tribe must have

greater than 97% sequence similarity, each clade must have greater than 95% sequence similarity, and lineages are monophyletic groups within phyla, typically with 85-90% sequence similarity. This taxonomy allows comparison of uncultured freshwater bacteria between lakes and research groups.

One idea that has gained popularity in recent years is that freshwater bacteria can be classified by their "lifestyles." It has long been noticed that bacteria can be roughly grouped by abundance patterns. Taxa can be abundant or rare in an ecosystem, and they can be persistent or transient. For example, abundant and transient microbes can be called "conditionally rare taxa" (33). In freshwater, toxic *Cyanobacteria* would be conditionally rare, although many non-toxic bacteria show this same trend. Presumably, the timing of blooms reflects some environmental driver in the lake. Another classification is by the range of carbon substrates consumed. Borrowing terms from niche ecology, a bacterium that can degrade many types of carbon is considered a generalist, while a picky bacterium would be a specialist. Growth rate and motility are two other traits that can be factored into lifestyle.

In freshwater, these traits were used to group bacterial taxa into four lifestyles: slow and augmented responders, fast and reduced responders, passive and streamlined, and vagabonds (34). The slow and augmented responders tend to have large genomes, a wide array of potential carbon substrates, and slow growth rates. Members of *Verrucomicrobia*, which are ubiquitous in freshwater and encode a variety of putative glycoside hydrolases, would fall into to this category (35). Fast and reduced responders

have tend to have genomes of intermediate size, high predicted growth rates, more specialized carbon degradation pathways, and motility. Presumably, the "bloomers" and conditionally rare taxa mentioned previously would meet this definition. Vagabonds are taxa that are found in other ecosystems and likely do not spend their entire life cycle in lakes; whether or not they are metabolically active in lakes is a matter for debate. Finally, the passive and streamlined bacteria have small genomes, slow predicted growth rates, few potential carbon sources, and little to no evidence of motility. Counterintuitively, members of this group are some of the most ubiquitous, abundant, and persistent bacteria in lakes. Examples of passive and streamlined freshwater bacteria include *Polynucleobacter necessarius,* which is commonly found in lakes and comprises 20% of bacterioplankton cells on average (36), and the *Actinobacteria* lineage acI, which can make up to 50% of bacterioplankton cells and is globally distributed (37).

While these lifestyle categories are not all-encompassing, they do provide a framework in which to place freshwater bacteria. One insight from these lifestyles is that cultured isolates from freshwater tend to be slow and augmented responders, such as members of *Rhodobacter* and *Chlorobiales (38, 39).* Meanwhile, the more abundant streamlined bacteria have proven more difficult to grow in the laboratory. Isolation of *P. necessarius* required acclimatization and the removal of potential competitors (40). While acI had previously been enriched using the dilution-to-extinction method (41), it was only recently isolated by adding catalase to the culture media (42). The success of

this approach was surprising, given that acI encodes and produces its own catalase. This result points to dependence on catalase production from another community member. Until a larger proportion of freshwater bacteria can be isolated, culture-independent methods are needed to prevent conclusions biased towards bacteria with large genomes and flexible metabolisms.

Time for time series

Time series analysis is a technique routinely employed in ecology. An ideal study would include replicate samples from identical lakes (43); however, finding lakes that are truly similar enough to be considered replicate observations of the same lake type is often impossible. For example, North and South Sparkling Bog were considered to be highly similar lakes based on their similar chemical gradients and geographic proximity in a 2012 study, where North Sparkling Bog was perturbed by mixing and South Sparkling Bog was used as a control (44). However, my analysis of a 16S rRNA amplicon time series in Chapter 3 showed significant differences in both richness and community composition between these two lakes (45). Time series can be used to assess variation within study sites and link that variation to environmental variables, and they are particularly valuable in situations such as lake sampling where true replicates and controls are not available (46). Moreover, time series may reveal trends and dynamics that cannot be observed in single timepoint studies (47). In this thesis, I use time series extensively to both identify trends over time and to assess variation within my study sites.

A meta-analysis of time series spanning one to three years found positive species-time relationships, indicating that more taxa are observed as the duration of sampling increases, either due to incomplete sampling, extinction and immigration, or speciation (48). In one freshwater lake, the amount of change in the bacterial community over a single day was equivalent to dissimilarity between sampling points ten meters apart (49). Conversely, bacterial communities can also change gradually over extremely long time scales, as they are sensitive to changes in environmental parameters such as nutrient availability and temperature. Wetland ecosystems and their carbon emissions are expected to change on scales greater than 300 years (50); as these emissions are the result of bacterial processes, we expect that the bacterial community will change on the same time scale as its ecosystem. Changes in marine phytoplankton regimes have been observed to occur over the past millennium, correlating with shifts in climate (51). With such a large range of potential time scales for change, we now recognize the need to more rigorously consider the duration and frequency of sampling in microbial ecology.

Multi-year studies of bacterial communities are less common due to their logistical difficulties and the need for stable funding, but results from the United States National Science Foundation funded Microbial Observatory and Long Term Ecological Research (LTER) projects are exemplary. 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 (52), while research at the

Sapelo Island – Microbial Observatory has led the field in integrating genomic data with environmental data (53). While there are several well-established long-term time series in marine systems, studies at this scale in freshwater are rare. 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 (54), 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 and are used extensively in this thesis. Chapter 3 details a multi-year time series of bacterial 16S rRNA gene amplicon data, allowing analysis of variation, seasonal trends, and community composition over time. Chapter 4 employs a metagenomic time series, paired with genomes recovered from that data using time series-resolved binning.Chapter 5 describes a short-term time series of metatranscriptomic data over two days, revealing interactions between taxa and informing the variability in future, long-term metatranscriptomic studies of freshwater. This time series is further used in Chapter 6, where aggregating metatranscriptomic data over two days provides a more detailed picture of gene expression than a single timepoint.

Freshwater in the "omics" era

The advent of next-generation sequencing transformed the field of microbiology, and freshwater microbial ecology is no exception. While sequencing is a powerful tool for observing thousands of microbial taxa at once, it presents unique drawbacks and

challenges. Every decision made in the sequencing process introduces bias, and analyzing sequencing data is often more difficult and time-consuming than producing it. However, the advantages provided by sequencing have made it the method of choice for many microbiome studies, and it is a method that I rely on extensively in this thesis (Table 1.1).

The first type of sequencing data I use is 16S rRNA gene amplicon sequencing. The 16S ribosomal RNA subunit is a gene found in all bacteria that contains both conserved and variable regions. This makes it an excellent target for phylogenetic studies. Prior to the sequencing era, the ribosomal RNA region was the marker for changes in community composition using Automated Ribosomal Intergenic Spacer Analysis (ARISA) (55). ARISA was extensively used in freshwater microbial communities to study responses to disturbance (44), to compare community changes across lakes (56), and to infer interactions between taxa (57). However, ARISA can only infer changes in community composition, not identify taxa in the community. 16S rRNA gene amplicon sequencing can be used to both classify sequences and measure relative abundance at a finer resolution than ARISA. It is important to note that sequencing provides relative abundance data, not absolute abundance data; if one taxon in the lake increases in abundance, the abundances of other taxa would appear to decrease in a 16S study. There are programs available to correct for the autocorrelations produced by this effect (58). Additionally, primer design and extraction method can impact the results of this type of sequencing (59, 60). Despite these limitations, 16S rRNA amplicon

sequencing is a powerful tool in a microbial ecologist's arsenal and a great candidate for time series analysis.

Metagenomics, used in Chapter 4, is the sequencing of all DNA in an environmental sample. This method has been used in freshwater to compare gene content across ecosystems or over time. For example, metagenomics has been used to predict the distribution of methylotrophs in freshwater (61), to identify carbon fixation pathways in humic lakes (62), and to study members of the candidate phyla radiation in groundwater (63). While metagenomics provides detailed information on the abundance of genes, classifying the genes in a metagenome can be difficult. One way to improve taxonomic classifications is to reconstruct genomes from metagenomes, known as "metagenome-assembled genomes" or MAGs. Although this approach aggregates genetically similar populations into often incomplete genomes, MAGs can be quite informative about the potential metabolic capabilities of uncultured taxa. The ubiquitous freshwater lineage acI has been studied using MAGs, which suggested additional nutrient sources that could be used by this taxon (64). Similarly, genes encoding diverse carbohydrate-active enzymes were identified in MAGs of freshwater Verrucomicrobia (35). While the presence of metabolic pathways inferred from gene content must be experimentally verified before being considered fact, analysis of MAGs can provide focus to research on complex communities.

Metatranscriptomics, the basis of Chapter 5, is the sequencing of environmental RNA and is one of the more recently applied sequencing techniques. The main challenge

in the application of metatranscriptomics is the RNA itself. We are interested in sequencing RNA because it degrades in minutes, providing a snapshot of the cell's metabolism in a short window of time (65). This also means that environmental RNA must be collected quickly and immediately immobilized, preferably by flash freezing samples in liquid nitrogen. Metatranscriptomes are also notoriously variable and require more replicates than metagenomes to ensure accurate conclusions. Add to this the fact that the human body is constantly shedding RNA-degrading enzymes as a defense against viruses, and collecting RNA in the field becomes a significant challenge. Metatranscriptomic studies of freshwater have been performed, but they are much less common than metagenomic studies, and often have smaller sample sizes (66). Metatranscriptomics does have its limitations, one key issue being the lack of evidence that mRNA expression correlates with the abundances of the encoded proteins in the cell (65). However, results from marine metatranscriptomics demonstrate that this method still has much to add to our understanding of freshwater microbial communities (53, 67).

Table 1.3. Sequencing methods used in this thesis. As each type of sequencing has unique strengths and weakness, a variety of methods were used in this thesis. The abbreviations, location in the thesis, and information about each sequencing method used is reported here.

Method	Abbreviation	Used in:	Strengths	Weaknesses
16S rRNA gene amplicon sequencing	16S sequencing	Chapter 3	- Provides information about taxonomy and abundance simultaneously - Low price and fast speed means many samples can be sequenced	- Abundance is relative, not absolute - Strain-level differences can be masked by similarity in the 16S region - Does not provide functional information
Metagenomics		Chapter 4, Chapter 5	- Provides information about gene content in an environment - Less biased than primer-based methods	- Linking function and taxonomy not straightforward - Gene annotations suggest, but do not prove, presence of a function
Metagenome-assembled genomes	MAGs	Chapter 4	- Improves gene-level classifications - Can be used to compare genomes across environments - Can infer metabolic pathways	- Genomes often incomplete or contaminated - Represents the average of a population - Gene annotations suggest, but do not prove, presence of metabolic pathways
Metatranscriptomics		Chapter 5	- Reveals which genes are actively transcribed in the environment - Co-expression can be used to strengthen pathway predictions	- Highly variable - Requires a reference database for classification and annotation of reads - Difficult to collect and process - ribosomal RNA must be degraded, which adds bias
Single amplified genomes	SAGs	Chapter 5	- Produces high quality genomes from individuals in the environment - Provides more information about genome structure and mobile genetic elements than MAGs	- Expensive and time-consuming - No abundance information unless combined with other methods

Single amplified genomes (SAGs) are the final type of sequencing used in this thesis. In this method, environmental cells are sorted by flow cytometry before being amplified and sequenced. While this process is time-consuming and expensive, SAGs have advantages over MAGs - they often contain additional circular DNA such as viruses and plasmids and represent a single strain rather than an average of a population (68). Functional screening can be added to the sorting step (69), or cells for input as SAGs can be chosen after sorting based on amplification of the 16S rRNA gene. SAGs are particularly powerful when used with other types of sequencing, as they can compensate for limitations in combination (70). SAGs are used in Chapters 5 as additional references for classification of metatranscriptomic reads, and they provide strain-level resolution of populations in Chapter 6.

Interactions between freshwater microbes

A major theme in this thesis is interactions within microbial communities. For much of the history of microbiology, microbes have been studied in isolation, a condition contrary to their natural state. We know that freshwater microbial communities are diverse and dynamic, but interactions between the microbes within those communities are not well-characterized. Connectivity and network properties derived from correlations between taxa abundances can shed light on potential interactions, but these analyses do not distinguish between taxa interactions and shared environmental drivers (71, 72). Still, a combination of correlative studies and

experimental studies suggest complex and highly relevant interactions between taxa in freshwater microbial communities.

As freshwater microbes form the base of the aquatic food web, predatory interactions significantly impact the community. The two main forms of predation on prokaryotic microbes are grazing by protists and viral infection. These top-down factors are associated with increased prokaryotic growth efficiency, presumably to compensate for higher mortality (73). Predatory interactions in these systems can be species-specific, with predators demonstrating a preference towards certain species of prokaryotes (74) and prey developing defense strategies such as filaments, small or large cell sizes, or high growth rates (75).

Interactions between phototrophic and heterotrophic microbes appear to be a driver of nutrient cycling and community composition in freshwater. Synchrony in community change between these two groups has been well documented (57, 76). Specifically, phototrophic microbes produce metabolites that are released extracellularly or through decay and then consumed by heterotrophic microbes. Heterotrophic bacteria with genes encoding enzymes to degrade glycolate, a photorespiration product, were observed to correlate with the phytoplankton community in humic lakes (77). Cultured isolates of the ubiquitous freshwater bacterial taxa *Limnohabitans* can grow using algal exudates as a substrate, and strains of *Limnohabitans* show preferences for exudates produced by different algal taxa (78). Metabolite exchange between phototrophs and

heterotrophs has been studied via transcriptomics in marine systems (67, 79); these results still need to be verified in freshwater systems.

From a microbiology perspective, the mechanism of interactions between aquatic microbes is an interesting research topic. In marine systems, extracellular vesicles carry chemical signals (and potentially viruses) between microbes (80). Electron transfer via cytochromes or conductive pili can be used to exchange energy over long distances and foster cooperation within microbial communities (81). Fundamental questions of cooperation, communication, community functions, and cheating can be addressed in freshwater microbial communities, while also shedding light on biogeochemical cycling.

Study sites in this thesis

Any boating or fishing enthusiast can tell you that there is a large amount of variation between lakes. Size, color, nutrient concentrations, and differences in the surrounding landscape are a few variables that can impact lake ecology, including the resident microbes. Comparing across lakes provides a natural contrast that can be used to investigate differences in the microbial communities, and it makes findings more generalizable to other bodies of freshwater. In this thesis, I focus on lakes near Madison and Minocqua, Wisconsin, representing three trophic statuses determined by carbon, nitrogen, and phosphorus concentrations: eutrophic, oligotrophic, and humic (Table 1.4). Many of these lakes are part of Microbial Observatory sites and the North Temperate Lakes - Long Term Ecological Research (NTL-LTER) program (82). These

pre-existing long-term datasets and continuing data collection, microbial and otherwise, are a valuable resource for my research.

The North Temperate Lakes - Microbial Observatory time series was collected from eight humic lakes near Minocqua in the boreal region of northern Wisconsin. Humic lakes contain high levels of dissolved organic carbon in the form of humic and fulvic acids, resulting in dark, acidic, "tea-colored" water. They are also commonly called bog lakes and occasionally "dystrophic" lakes. Due to their dark color, bog lakes absorb heat from sunlight, resulting in strong stratification during the summer. The top layer in a stratified lake, called the "epilimnion," is oxygen-rich and warm. At the lake bottom, a cold layer called the "hypolimnion" is formed, becoming anoxic almost immediately in darkly stained bog lakes. The border between these two regions is called the "thermocline." The transitions between mixing of these two layers and stratification occur rapidly in these systems, and at different frequencies (called mixing regimes) depending on the depth, surface area, and wind exposure of the lake. Changes in bacterial community composition along the vertical gradients established during stratification are well documented (83, 84). Mixing has been shown to be a disturbance to the bacterial communities in bog lakes (44). The bacterial communities in bog lakes are still being characterized, but contain both ubiquitous freshwater organisms (36, 41) and members of the candidate phyla radiation (85). Seasonality in freshwater lakes is thought to be the norm rather than the exception (86, 87); however, multiple years of

sampling are needed to confirm these prior findings. Two of the bog lakes (Trout Bog and Crystal Bog) are also NTL-LTER core sites.

Lake Mendota, a eutrophic lake located in Madison, WI, is arguably one of the best studied lakes in the world (88). This relatively large lake is surrounded by an urban and agricultural watersheds, which lead to increased nitrogen and phosphorus inputs and high levels of primary production. Like bog lakes, Lake Mendota stratifies seasonally into an epilimnion and a hypolimnion. However, Lake Mendota's larger surface area and greater water clarity results in more wind-induced mixing events and variation in the location of the thermocline. Because of its proximity to the city of Madison, Lake Mendota is strongly impacted by human activity and affects humans in return. Toxic cyanobacterial blooms occur regularly on this lake, and it has recently had two invasive species introductions: the spiny water flea in 2009 and the zebra mussel in 2015. Lake Mendota is the subject of its own decadal Microbial Observatory project in addition to being a core site for the NTL-LTER. Recurring annual trends in microbial communities have been observed in Lake Mendota through both ARISA (89) and 16S rRNA gene amplicon sequencing (90).

Table 1.4. Study sites used in this thesis. Lakes used in this body of research were chosen to include a variety of trophic statuses and mixing regimes. Lakes that are sites of the North Temperate Lakes - Long Term Ecological Research program (NTL-LTER) were preferentially included, as extensive environmental data is available for these sites. Many of these lakes have been used in previous microbial ecology research. Additional limnological data is available in the thesis chapters describing the microbiology of these lakes.

Lake	Surface Area (m^2)	Depth (m)	Nearest town	Trophic status	Mixing regime	LTER site?
Lake Mendota	39,610,000	25.3	Madison	Eutrophic	Dimictic	Yes
Sparkling Lake	640,000	20	Arbor Vitae	Oligotrophic	Dimictic	Yes
Trout Bog	10,100	7	Boulder Junction	Humic	Dimictic	Yes
Crystal Bog	5,600	2.5	Boulder Junction	Humic	Polymictic	Yes
Mary Lake	12,000	21.5	Winchester	Humic	Meromictic	No
Forestry Bog	1,300	2	Boulder Junction	Humic	Polymictic	No
West Sparkling Bog	11,900	4.6	Arbor Vitae	Humic	Polymictic	No
North Sparkling Bog	4,700	4.5	Arbor Vitae	Humic	Dimictic	No
South Sparkling Bog	4,400	8	Arbor Vitae	Humic	Dimictic	No
Hell's Kitchen	30,000	19.3	Presque Isle	Humic	Meromictic	No

Sparkling Lake, representing the oligotrophic lakes, is located near Minocqua, WI, and is the least extensively studied of the lakes included in this work. While still a core NTL-LTER site, it is not a Microbial Observatory site. Bacterial genomes were sequenced from this lake in 2009 and have been used in studies examining the distribution of rhodopsins in freshwater (91) and the differences in subgroups of the ubiquitous freshwater lineage acI (92). Addition of samples from this lake to those from our two Microbial Observatory programs provides an excellent contrast in nutrient status.

In this thesis, I study nutrient cycling in freshwater microbial communities to advance our understanding of microbial ecology principles and of nutrient cycling in freshwater. The rich background of research on both freshwater microbiology and freshwater ecology, as well as the proliferation of newly applicable methods, make this a timely topic. I add to our knowledge in this area by performing next-generation sequencing techniques on freshwater microbial communities, an approach that provides broad insights on community composition, community dynamics, and potential metabolic functions.

Chapter 2: Objectives

The broad goal of my thesis is to apply time series sequencing methods to freshwater bacterial communities to investigate bacterial nutrient cycling in freshwater. The advent of next-generation sequencing methods, paired with the existing long-term datasets, has the potential to significantly improve our understanding of the dynamics and functions of freshwater bacteria.

Aim 1: Long-term dynamics of bacterial communities in bog lakes

Before we can begin to study the role of bacteria in freshwater nutrient cycling, the variability, community composition, and seasonality of freshwater bacterial communities assessed. While previous studies have investigated community dynamics using methods such as ARISA, clone libraries, or T-RFLP, these methods cannot simultaneously classify and quantify bacterial taxa. To assess how bacterial communities change over multi-year time scales, DNA samples were collected from bog lakes over five years during the ice-free period. These samples were submitted for 16S rRNA ribosomal gene amplicon sequencing through the Earth Microbiome Project (93), producing a time series of community data. In Chapter 3, I analyze community membership, seasonality, and differences between lakes to inform future experiments and analyses. This is research is the subject of the publication Linz et al., “Bacterial

community composition and dynamics spanning five years in freshwater bog lakes." *mSphere*, 2017 (45).

Aim 2: Inferring metabolic traits from metagenome-assembled genomes

Because of the complexity of freshwater bacterial communities demonstrated in Aim 1, culturing the thousands of populations in lakes and studying their metabolisms in the laboratory would be a herculean task. Instead, we used our metagenomic time series and metagenome-assembled genomes (MAGs) to predict metabolic pathways based on gene content. MAGs were recovered from two of our long term study sites, Lake Mendota and Trout Bog, which having contrasting chemical characteristics. We expected to find different metabolic functions suggested by the gene content in metagenomes and genomes from each lake. In Chapter 4, I describe the differences and similarities observed in metagenomic gene content between our two study sites, as well as additional information about pathways and taxonomy gained from analyzing MAGs.

Aim 3: Diel metatranscriptomics of eutrophic, oligotrophic, and humic lakes

After determining which bacteria are present in freshwater and what genes and potential metabolic capabilities they possess, the logical next step is to learn which bacteria and genes are active. Research from marine systems suggests that gene expression can vary on a 24-hour cycle, and that this trend is driven by phototrophic microbes. Given that freshwater phototrophic and heterotrophic microbes have been

shown to exchange metabolites, we collected RNA samples every four hours for 48 hours from Lake Mendota, Trout Bog, and Sparkling Lake to identify which taxa co-express and what metabolites they are expressing. In Chapter 5, I compare gene expression in day versus night and find that genes related to primary production are upregulated in daylight, while genes related to sugar consumption are upregulated at night. This suggests that there is an exchange of carbon between the phototrophic and heterotrophic communities.

Aim 4: Niche partitioning in closely related bacteria.

A major question in microbial ecology is how bacterial speciation occurs. Highly genetically similar populations have been observed co-existing in freshwater, suggesting that they have some manner of niche partitioning not revealed through gene content. In Chapter 6, I use the two day metatranscriptomic time series to compare expression in these closely related populations. I identify transporters not expressed in all of these populations, suggesting that niche partitioning is highly dynamic and can occur via transcriptional regulation.

Projects not covered in this thesis

Science in its published form often appears as a progression of clear and logical steps; however, these finished products often do not reflect the many dead-ends, side projects, or failed experiments that occur in the course of a PhD. Below is a list of research projects not discussed later in this thesis.

Isolating acI. With the help of a post-doctoral researcher, I enriched acI from Trout Bog. However, the enrichments did not maintain their density indefinitely, and attempts to inoculate the enrichments in an artificial lake water media failed. Another research group recently found that catalase addition is needed to grow acI in artificial media.

BrdU incorporation. Bromo-deoxyuridine is a nucleotide homolog that can be incorporated into actively replicating DNA and precipitated using a targeted antibody. This approach could theoretically be used to sequence metagenomes of only actively replicating organisms, but the yield of BrdU-tagged DNA is too low to sequence. I optimized the protocol to increase yield from lake water by 10-fold, but still could not produce enough labelled DNA for sequencing. A concurrent project at the Joint Genome Institute optimizing the same technique on liters of batch *E. coli* cultures also do not provide enough yield.

Alum addition. To reduce phosphorus inputs to area lakes, the City of Madison ran a pilot program to precipitate phosphorus and other compounds from stormwater using alum before it could enter the water system. One of the test sites was in the UW-Madison Arboretum, and we were recruited to assess the ecological impact of this manipulation on microbial communities. We saw no significant changes in richness or seasonal dynamics during the pilot program, and alum addition was ultimately unsuccessful in effectively reducing phosphorus levels in Madison lakes.

Candidate phyla in Mary Lake. I used depth discrete metagenomes and 16S rRNA amplicon data from Mary Lake to demonstrate that ultra-small members of the

candidate phyla radiation (CPR) contributed nearly 5% of microbial abundance in this site and were significantly associated with other, non-CPR microbes. I wrote an NSF DDIG proposal to further investigate freshwater CPR microbes by tracking isotopically labeled carbon substrates in ultra-small bacteria using Raman microspectroscopy in collaboration with the Joint Genome Institute, but this proposal was not funded.

Eutrophication and bacterial growth efficiency. Bacterial growth efficiency (BGE) is the ratio of carbon incorporated as biomass to carbon respired. In my preliminary exam proposal, Aim 3 was to experimentally investigate how eutrophication impacts BGE, as studies comparing BGE in sites with different levels of eutrophication produced conflicting results. With the help of an undergraduate researcher, I set up mesocosms of oligotrophic lake water, added plant fertilizer, and calculated BGE over time. However, high levels of variability in BGE between replicate mesocosms prevented further use of this data and further experiments. Likely, substantially larger mesocosms would have been needed to mitigate bottle effects.

Bacterial fuel cells. After one of our lab members identified genes potentially encoding extracellular electron transfer, we built small-scale fuel cells to confirm that this process can occur in the water column of Trout Bog. These fuel cells were highly successful, as was an *in situ* setup suspended from a buoy in Trout Bog. Extracellular electron transfer will be the subject of future research in the McMahon Lab.

Chapter 3: Bacterial community composition and dynamics in bog lakes

Alexandra M. Linz[1], Benjamin C. Crary[1], Ashley Shade[2], Sarah Owens[3], Jack A. Gilbert[3,4,5], Rob Knight[6,7,8], and Katherine D. McMahon[1,9]

[1]Department of Bacteriology, University of Wisconsin – Madison, Madison, WI, 53706, USA, [2]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, 48824, USA, [3]Biosciences Division, Argonne National Laboratory, Argonne, IL 60439, USA, [4]Computation Institute, University of Chicago, Chicago, IL 60637, USA, [5]Department of Ecology and Evolution, Department of Surgery, University of Chicago, Chicago, IL 60637, [6]Center for Microbiome Innovation, Jacobs School of Engineering, University of California, San Diego, La Jolla, California 92093, USA., [7]Department of Pediatrics, University of California, San Diego School of Medicine, La Jolla, California 92093, USA., [8]Department of Computer Science and Engineering, Jacobs School of Engineering, University of California San Diego, La Jolla, California 92093, USA., [9]Department of Civil and Environmental Engineering, University of Wisconsin – Madison, Madison, WI, 53706

A.M. Linz performed analyses and wrote manuscript
B.C. Crary performed analyses and edited manuscript
A. Shade designed and performed experiments and edited manuscript
S. Owens performed experiments
J.A. Gilbert contributed funding and edited manuscript
R. Knight contributed funding and edited manuscript
K.D. McMahon designed experiments, contributed funding, and wrote manuscript

Publication: Linz, A. M., Crary, B. C., Shade, A., Owens, S., Gilbert, J. A., Knight, R., & McMahon, K. D. (2017). Bacterial community composition and dynamics spanning five years in freshwater bog lakes. *mSphere*, *2*(3), e00169-17.

Introduction

One of the major goals of microbial ecology is to predict bacterial community composition. However, we have only a cursory knowledge of the factors that would allow us to predict bacterial community dynamics. To characterize the diversity and dynamics of an ecosystem's bacterial community, sampling the same site multiple times is just as necessary as sampling replicate ecosystems. Additionally, the sampling frequency must match the rate of change of the process being studied. We must first understand the scales on which bacterial communities change before we can design experiments that capture a full range of natural variation. To assess bacterial community composition and dynamics on long time scales in our bog lake study sites (Table 3.1), we built and analyzed a multi-year time series of 16S rRNA gene amplicon data from two depths in eight lakes.

The bog lakes in this study have been model systems for freshwater microbial ecology for many years. Early studies used Automated Ribosomal Intergenic Spacer Analysis (ARISA), a fingerprinting technique for identifying unique bacterial taxa in environmental samples (55). Our research built upon these studies and added information about the taxonomic identities of bacterial groups. For example, persistent and unique bacterial groups were detected in the bog lakes using ARISA (94). Differences in richness and community membership were previously detected within one year, between Crystal Bog, Trout Bog, and Mary Lake, three sites representative of the three mixing regime categories of polymictic, dimictic, and meromictic (95).

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

Table 3.1. Study sites in the North Temperate Lakes Microbial Observatory. The lakes included in this time series are small, humic bog lakes in the boreal region near Minocqua, Wisconsin, USA. They range in depth from 2 to 21.5 meters and encompass a range of water column mixing frequencies (termed regimes). Dimictic lakes mix twice per year, typically in fall and spring, while polymictic lakes can mix more than twice throughout the spring, summer, and fall. Meromictic lakes have no recorded mixing events. pH was measured in 2007, while nutrient data was measured in 2008 (with the exceptions of FB, WS, and HK, measured in 2007); both measurements were taken concurrently with the bacterial biomass collection from the same water sample. Standard deviations for DOC/DIC are reported in parentheses. When two values are present in a single box, the first represents the epilimnion value and the second represents the hypolimnion value.

	Forestry Bog	**Crystal Bog**	**North Sparkling Bog**	**West Sparkling Bog**	**Trout Bog**	**South Sparkling Bog**	**Hell's Kitchen**	**Mary Lake**
ID	FB	CB	NS	WS	TB	SS	HK	MA
Depth (m)	2	2.5	4.5	4.6	7	8	19.3	21.5
Surface area (m^2)	1300	5600	4700	11900	10100	4400	30000	12000
Approx. Volume (m^3)	867	4667	7050	18247	23567	11733	193000	86000
Mixing regime	Polymictic	Polymictic	Dimictic	Dimictic	Dimictic	Dimictic	Meromictic	Meromictic
GPS coordinates	46.04777, -89.651248	46.007639, -89.606341	46.004819, -89.705214	46.004633, -89.709082	46.041140, -89.686352	46.003334, -89.705296	46.186674, -89.702510	46.250764, -89.900419
Years sampled	2007	2007, 2009	2007, 2008, 2009	2007	2005, 2007, 2008, 2009	2007, 2008, 2009	2007	2005, 2007, 2008, 2009
pH	4.97, 4.85	4.49, 4.41	4.69, 4.80	5.22, 5.14	4.60, 4.78	4.46, 4.94		5.81, 5.72
Dissolved inorganic carbon (ppm)	0.94, 1.46	0.69, 1.72	1.12, 2.31	0.76, 1.56	1.73, 4.47	1.97, 6.42	2.91, 9.70	5.54, 12.38
Std Dev	(0.28, 1.17)	(0.15, 0.50)	(0.23, 0.72)	(0.17, 0.36)	(0.66, 54)	(0.24, 1.56)	(0.35, 1.03)	(5.66, 7.69)
Dissolved organic carbon (ppm)	10.22, 8.96	15.47, 13.6	10.05, 10.40	7.26, 7.27	19.87, 20.58	12.40, 21.92	7.26, 7.33	20.63, 67.10
Std Dev	(0.59, 0.10)	(4.12, 0.82)	(1.16, 0.96)	(0.43, 0.73)	(2.76, 1.17)	(0.38, 4.76)	(1.03, 0.12)	(1.91, 72.67)
Total nitrogen (ppb)		620.57, 846.00	629.09, 809.45		737.71, 1121.00	813.88, 1498		1332.57, 3652.38
Total phosphorus (ppb)		30.00, 38.86	78.00, 135.45		50.57, 53.25	48.63, 69.14		78.00, 303.50
Total dissolved nitrogen (ppb)		1290.19, 490.13	442.39, 586.56		582.5, 820.21	451.63, 1179.21		1024.5, 3220.14
Total dissolved phosphorus (ppb)		84.25, 14.88	70.22, 22.67		34.5, 31.57	16.25, 18.29		71.13, 228

Methods

Sample collection

Water was collected from eight bog lakes during the summers of 2005, 2007, 2008 and 2009, as previously described (95). Sampling occurred at approximately weekly intervals and primarily during the summer stratified period (May – Aug) (Figure 3.1). Sites were not sampled continuously over the entire time series, and metadata is available only for a subset of samples. 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, two biological replicates were taken by filtering approximately 150 mL from each well-mixed sample through 0.22 micron polyethersulfone filters (Supor, 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 (96).

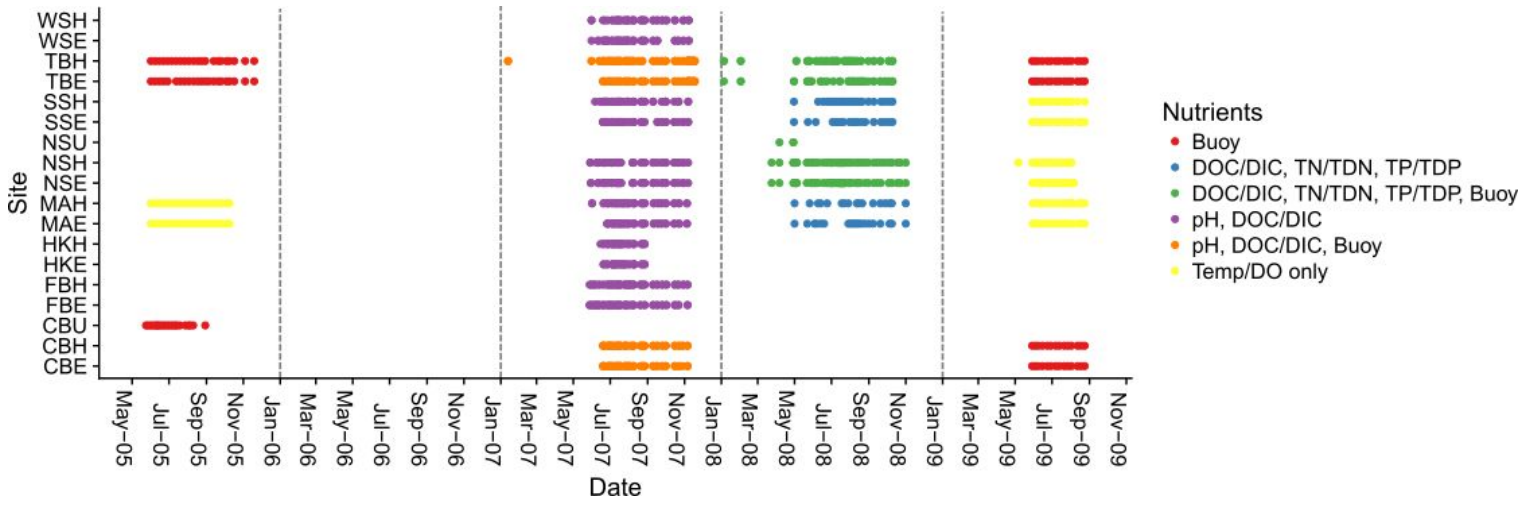

Figure 3.1. Sampling frequency and associated environmental data for the NTL Microbial Observatory. Eight bog lakes were sampled at two depths during the ice-free seasons of 2005, 2007, 2008, and 2009. A temperature and dissolved oxygen (DO) profile was measured with every sample. Some years and lakes contain additional chemical data such as pH, dissolved organic carbon (DOC), dissolved inorganic carbon (DIC), total nitrogen (TN), total dissolved nitrogen (TDN), total phosphorus (TP), and total dissolved phosphorus (TDP). Trout Bog (TB), Crystal Bog (CB), and North Sparkling Bog (NS) hosted an NTL-LTER autonomous data collection buoy during all or part of the microbial sampling period.

The sampling sites are located near Boulder Junction, WI, and were chosen to include lakes represent the three mixing regimes of polymictic (multiple mixing events per year), dimictic (two mixing events per year, usually in spring and fall), and meromictic (no recorded mixing events) (Table 1). Trout Bog and Crystal Bog are also primary study sites for the North Temperate Lakes - Long Term Ecological Research Program (NTL-LTER), which measures a suite of chemical limnology parameters fortnightly during the open water season. The NTL-LTER also maintains autonomous sensing buoys on Trout Bog and Crystal Bog, allowing for more refined mixing event detection based on thermistor chain measurements.

Sequencing

A total of 1,510 DNA samples, including 547 biological replicates, were sequenced by the Earth Microbiome Project according to their standard protocols in 2010, using the original V4 primers (FWD:GTGCCAGCMGCCGCGGTAA; REV:GGACTACHVGGGTWTCTAAT) (97). 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 (98). 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 (99). Thirty-three chimeras and 340 chloroplast sequences (based on pre-clustering and classification with the Greengenes 16S rRNA gene database, May 2013) (100) were removed. Samples were rarefied to 2,500 reads; samples with less than 2,500 reads were omitted, resulting in 1,387 remaining samples. The rarefaction cutoff used was determined based on the results of simulation; 2,500 reads was chosen to maximize the number of samples

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

Representative sequences for each OTU were classified in either our curated freshwater database (32) or the Greengenes database based on the output of NCBI-BLAST (blast+ 2.2.3.1) (101). 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 (102), 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) (103).

Statistics

Statistical analysis was performed in R v3.3.2 (R Development Core Team, 2008. R: A language and environment for statistical computing.). Significant differences in richness between lakes was tested using a pairwise Wilcoxon sum rank test with a Bonferroni adjustment in the R package “exactRankTests” (T. Hothorn and K. Hornik, 2015. exactRankTests: Exact Distributions for Rank and Permutation Tests). Similarity between samples was compared using weighted UniFrac distance, implemented in “phyloseq” (104). Weighted UniFrac distance was chosen because it explained the

greatest amount of variation in the first two axes of a principal coordinates analysis, performed in “vegan” (J. Oksanen, 2016. vegan: Community Ecology Package). Other metrics tested included unweighted UniFrac distance, Bray-Curtis Dissimilarity, and Jaccard Similarity; the output of all metrics were correlated. Significant clustering by year in PCoA and in dispersion between lakes was tested using PERMANOVA with the function adonis() in “vegan.” Trimming of rare taxa did not impact the clustering observed in ordinations, such as those present in Figure 2, even when taxa observed less than 1000 times were removed.

Indicator species analysis was performed using “indicspecies” (105). Only taxa with read abundances of at least 500 reads in 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” (H. Wickham, 2009. ggplot2: Elegant Graphics for Data Analysis) and “cowplot” (C. Wilke, 2016. cowplot: Streamlined Plot Themes and Plot Annotations for ‘ggplot2’). “reshape2” was used for data formatting (H. Wickham, 2007. Reshaping Data with the reshape Package).

Data availability

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

Results

Overview of community composition

We used a time series of 16S rRNA gene amplicon data to investigate bacterial community composition over time and across lakes. A total of 8,795 OTUs were detected in 1,387 samples. As is typical for most freshwater ecosystems, Proteobacteria, Actinobacteria, Bacteroidetes, and Verrucomicrobia were the most abundant phyla (Figure 3.2). Within these phyla, OTU abundance was highly uneven. Much of the abundance of Proteobacteria could be attributed to OTUs belonging to the well-known freshwater groups Polynucleobacter and Limnohabitans, and the freshwater lineage acI contributed disproportionately to the observed abundance of Actinobacteria. Like many microbial communities, unevenness was a recurring theme in this dataset, which had a long rare tail of OTUs and trends driven largely by the most abundant OTUs (106, 107).

These results show that the composition of our dataset is consistent with results from other bog lakes (84, 85).

Community richness

We hypothesized that water column mixing frequency was associated with alpha diversity. Observed richness was calculated for every sample at the OTU level, and samples were aggregated by lake and layer. Hypolimnia were generally richer than epilimnia (Figure 3.3). Significant differences in richness between lakes were detected using the Wilcoxon signed rank test with a Bonferroni correction for multiple pairwise comparisons (Table 3.2). For both layers, polymictic lakes had the fewest taxa, dimictic lakes had intermediate numbers of taxa, and meromictic lakes had the most taxa. This dataset includes two fall mixing events (Trout Bog 2007 and North Sparkling Bog 2008), as well as the artificial mixing event in North Sparkling Bog 2008 (44). Richness decreased sharply in mixed samples compared to those taken during the summer stratified period (Figure 3.4). The observed association between mixing frequency and richness suggests that water column mixing (or one or more co-varying environmental parameters) structures the bacterial community.

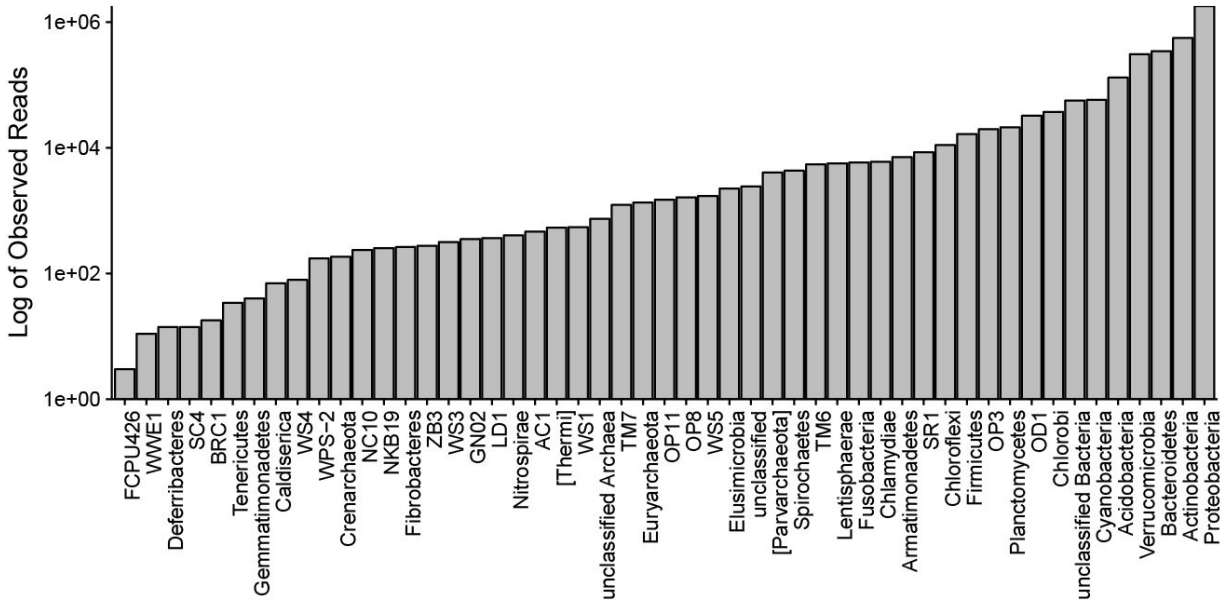

Figure 3.2. Phyla abundances summed across the 16S dataset. Proteobacteria, Actinobacteria, and Bacteroidetes were the most abundant phyla detected. Within each phylum, reads often belonged most frequently to a small number of lineages. Results are consistent with other published reports from humic bog lakes. Note that members of the candidate phyla radiation (OD1, SP3, SR1, etc.) are named based on the state of taxonomy in Greengenes version 13.5 and may have been given new names in more recent literature.

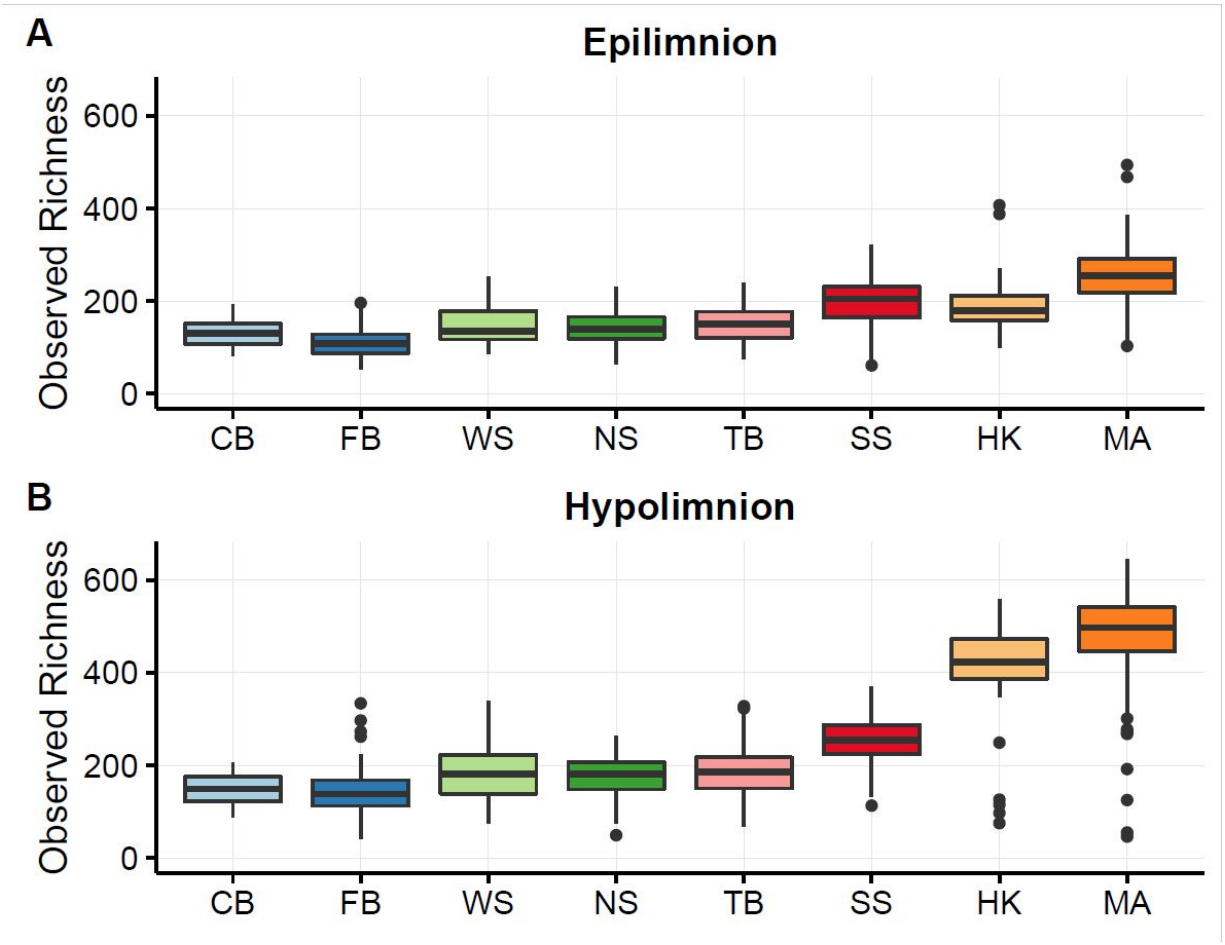

Figure 3.3. Richness by layer and lake. Lakes on the x axis are arranged by depth (see Table 1 for lake abbreviations and depth measurements). Significant differences between lakes were observed, as reported in Table 3.2.

Table 3.2. Significant differences in richness between lakes. Significance was tested using a Wilcoxon signed rank test with a Bonferroni correction for multiple pairwise comparisons. P-values resulting from this test are presented in this table. Significant ($p < 0.05$) values are highlighted. Table 2A shows the results from epilimnia, while Table 2B shows results from hypolimnia.

A	FBE	WSE	NSE	TBE	SSE	HKE	MAE
CBE	0.01	0.84	0.12	0.01	0.00	0.00	0.00
FBE		0.00	0.00	0.00	0.00	0.00	0.00
WSE			1.00	1.00	0.00	0.01	0.00
NSE				1.00	0.00	0.00	0.00
TBE					0.00	0.00	0.00
SSE						1.00	0.00
HKE							0.00

B	FBH	WSH	NSH	TBH	SSH	HKH	MAH
CBH	1.00	0.04	0.00	0.00	0.00	0.00	0.00
FBH		0.03	0.00	0.00	0.00	0.00	0.00
WSH			1.00	1.00	0.00	0.00	0.00
NSH				1.00	0.00	0.00	0.00
TBH					0.00	0.00	0.00
SSH						0.00	0.00
HKH							0.00

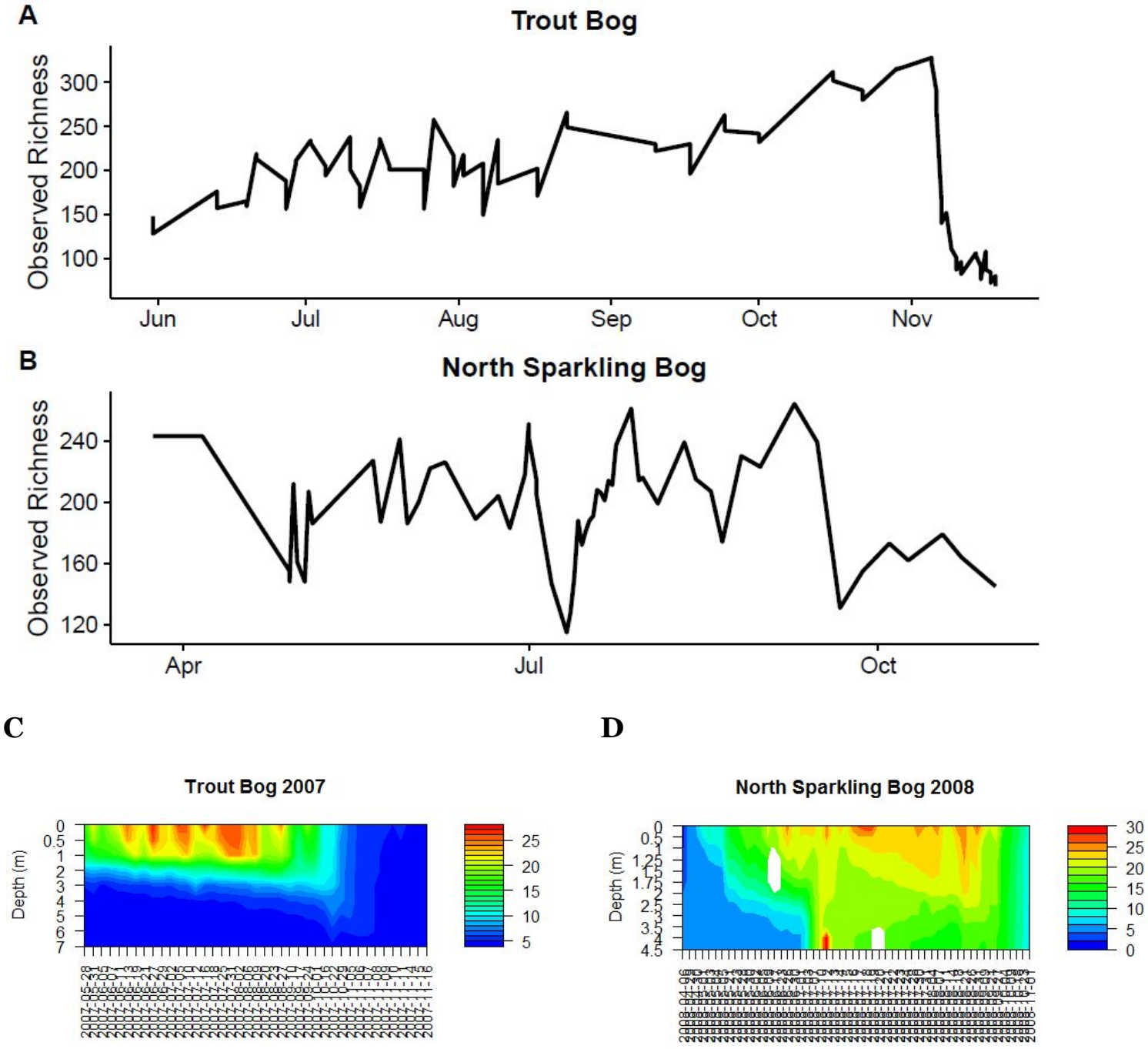

Figure 3.4. Richness over time. In addition to differences between lakes and layers, richness also varied over time. In Trout Bog and North Sparkling Bog (A, B), two lakes with captured mixing events, richness decreased when the water column was more evenly mixed (C, D). This suggests that mixing acts as a disturbance to the freshwater microbial communities.

Clusters of community composition

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

Within layers, mixing regime also explained differences in community composition (Figure 3.5). Clustering by mixing regime was significant by PERMANOVA in both epilimnia and hypolimnia samples ($r2 = 0.20$ and $r2 = 0.22$, respectively, and $p = 0.001$ in both groups). Site was a strong factor explaining community composition, with significant clustering in epilimnia ($p = 0.001$, $r2 = 0.34$) and hypolimnia ($p = 0.001$, $r2 = 0.49$). Date and mean water temperature did not describe the observed clustering as well as lake or mixing regime. Because PCoA can be susceptible to artifacts, we also performed a comparison of beta diversity between sites using a Bray-Curtis dissimilarity distance matrix without ordination; the same results were obtained. These findings demonstrate that thermal layer, lake, and mixing frequency are associated with changes in bacterial community composition.

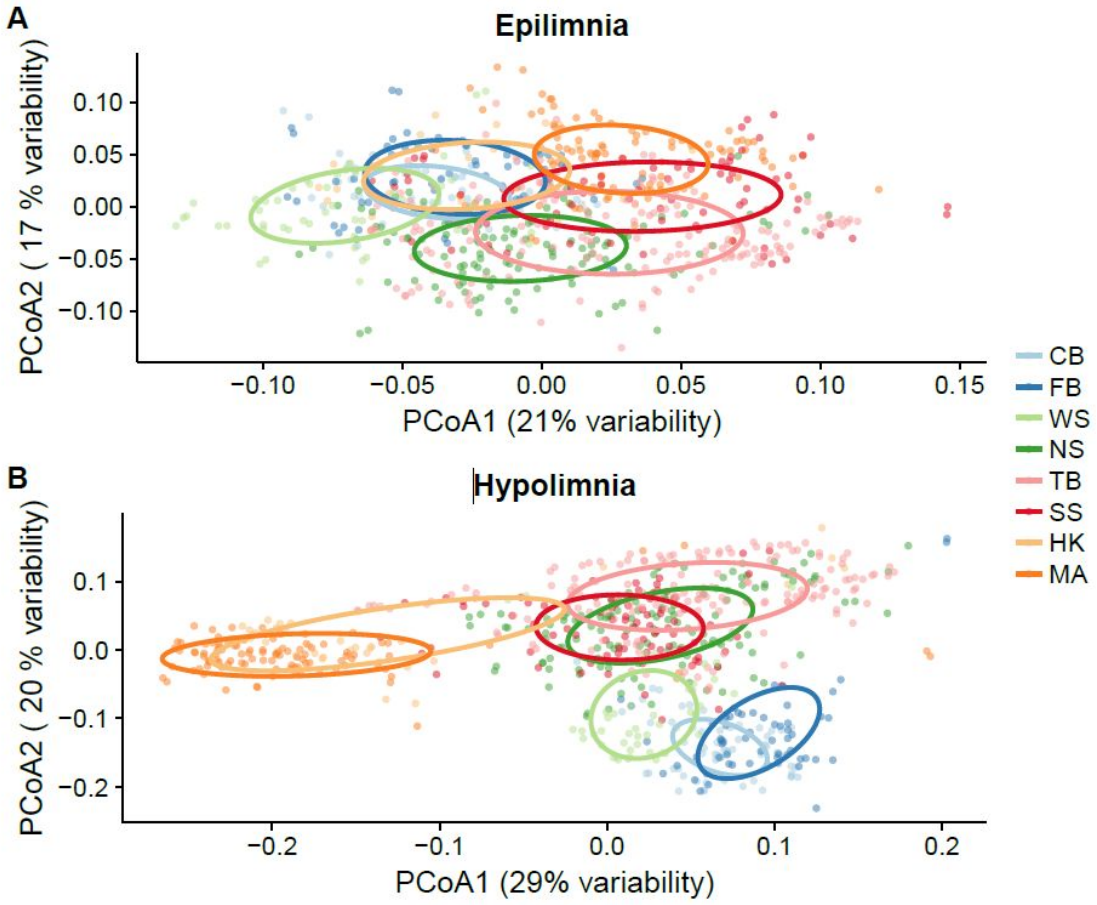

Figure 3.5. Principal coordinates analysis of lakes by layer. Samples collected from epilimnia (A) and hypolimnia (B) cluster significantly by lake and by mixing regime. However, this separation is more pronounced in hypolimnia than in epilimnia.

Variability and dispersion

While OTU-based community composition was distinct by layer, lake, and mixing regime, there was still variability over time. We used weighted UniFrac distance to quantify variability in beta diversity between samples within the same site and year.

Each year in each lake had a significantly different community composition, indicating interannual variability in the community composition (Figure 3.6a-c, Table 3.3). As multiple environmental variables changed in each year of sampling, it is not clear which (if any) could explain the observed annual shifts in community composition. We found no evidence of repeating seasonal trends during the stratified summer months in these lakes in time decay plots using weighted UniFrac distance. Likewise, we examined trends in the most abundant individual OTUs and did not observed repeatable annual trends, even when abundances in each year were normalized using z-scores.

Variability can also be assessed by measuring the beta diversity within a single site. We measured pairwise weighted UniFrac distance between samples in each each lake-layer (Figure 3.6d). This analysis showed that layers had significantly different levels of beta diversity within a single site for West Sparkling Bog, North Sparkling Bog, Trout Bog, South Sparkling Bog, and Mary Lake, as determined using a Wilcoxon signed rank test with a Bonferroni correction for multiple pairwise comparisons (Table 3.3). Within-site beta diversity was not significantly different in Crystal Bog, Forestry Bog, and Hell's Kitchen. Mean pairwise UniFrac distance was lower in the epilimnion than the hypolimnion in the West and North Sparkling Bogs, but higher in the other three significant sites. Performing the same analysis on a single year of data with approximately even numbers of samples from each site showed the same trends. This shows that the amount of variability in the bacterial community differs by site as well as by year.

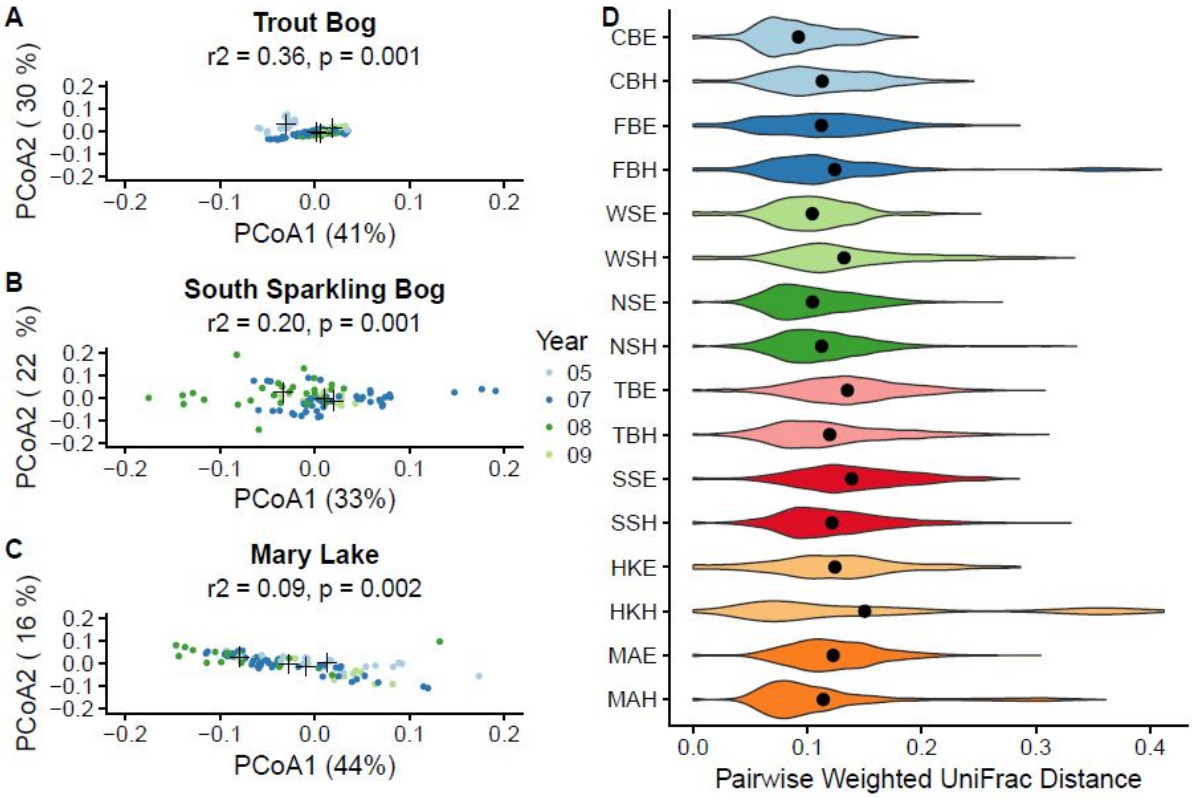

Figure 3.6. Principal coordinates analysis shows clustering by year within lakes and differences in dispersion between layers. Principal coordinates analysis using weighted UniFrac as the distance metric was used to measure the amount of interannual variation in the three lake hypolimnia with the longest time series (A-C). Black crosses indicated the centroid for each year. All hypolimnia showed significant clustering by year by PERMANOVA. Six outliers in Mary Lake from 2007 are not shown, as their coordinates lie outside the range specified for consistency between plots; these points were included in the PERMANOVA significance test. Panel D shows pairwise weighted UniFrac distance within each lake and layer including all samples. Lakes with significant differences between layers at $p < 0.05$ by a Wilcoxon signed rank test with a Bonferroni correction include Crystal Bog (CB), Forestry Bog (FB), North Sparkling Bog (NS), Trout Bog (TB), South Sparkling Bog (SS), and Mary Lake (MA).

Table 3.3. Significance of clustering by layer and year in Figure 3.6. PERMANOVA was used to determine if there was significant clustering in the principal coordinates analyses performed various subsets of the data. All subsets tested showed significant clustering, indicating that there are detectable differences in samples from different layers and years.

		Degrees Freedom	Sums of Squares	Mean Squares	F Statistic	Partial r^2	p-value
Epilimnion	Lakes	7	2.68	0.38	45.85	0.33	0.001***
	Residuals	663	5.53	0.01		0.67	
	Total	670	8.20			1.00	
	Regime	2	1.52	0.76	75.85	0.19	0.001***
	Residuals	668	6.68	0.01		0.81	
	Total	670	8.20			1.00	
Hypolimnion	Lakes	7	5.25	0.75	98.38	0.50	0.001***
	Residuals	682	5.20	0.01		0.50	
	Total	689	10.45			1.00	
	Regime	2	3.89	1.94	203.49	0.37	0.001***
	Residuals	687	6.56	0.01		0.63	
	Total	689	10.45			1.00	
Trout Bog	Year	3	0.30	0.10	30.25	0.36	0.001***
	Residuals	162	0.53	0.00		0.64	
	Total	165	0.83			1.00	
South Sparkling Bog	Year	2	0.11	0.06	10.57	0.20	0.001***
	Residuals	82	0.44	0.01		0.80	
	Total	84	0.55			1.00	
Mary Lake	Year	3	0.35	0.12	3.79	0.10	0.001***
	Residuals	99	3.04	0.03		0.90	
	Total	102	3.39			1.00	

The core community of bog lakes

One of the goals of this study was to determine the core bacterial community of bog lakes, and to determine if mixing regime affects core community membership. Our previous analyses showed that community composition was distinct in each layer and lake, while marked variability was observed within the same lake and layer. This prompted us to ask whether we had adequately sampled through time and space to fully

census the lakes. Still, rarefaction curves generated for the entire dataset and for each layer begin to level off, suggesting that we have indeed sampled the majority of taxa found in our study sites. To identify the taxa that comprise the bog lake core community, we defined "core" as being present in 90% of a group of samples, regardless of abundance in the fully curated dataset. Core taxa are reported as OTU number and taxonomic classification our freshwater-specific database (32). Four OTUs met this criteria for all samples in the full dataset: OTU0076 (bacI-A1), OTU0097 (PnecC), OTU0813 (acI-B2), and OTU0678 (LD28). These taxa were therefore also core to both epilimnia and hypolimnia. Additional taxa core to epilimnia also included OTU0004 (betI), OTU0184 (acI-B3), OTU0472 (Lhab-A4), and OTU0522 (alfI-A1), while additional hypolimnia core taxa included OTU0042 (Rhodo), OTU0053 (unclassified *Verrucomicrobia*), and OTU0189 (acI-B2).

We performed the same core analysis after combining OTUs assigned to the same tribe (previously defined as sharing ≥ 97% nucleotide identity in the nearly full length 16S rRNA gene and according to phylogenetic branch structure) into new groups. This revealed that certain tribes were core to the entire dataset or thermal layer even though their member OTUs were specific to certain sites. Notably, some OTUs were endemic to specific lakes, even though their corresponding tribe was found in multiple lakes/layers. OTUs not classified at the tribe level were not included. Results were similar to those observed at the OTU level, but yielded more core taxa. Tribes core to all samples included bacI-A1, PnecC, acI-B2, and LD28, but also betIII-A1 and acI-B4. In epilimnia,

the core tribes were bacI-A1, PnecC, betIII-A1, acI-B3, acI-B2, Lhab-A4, alfI-A1, LD28, and acI-B4, while in hypolimnia, they were Rhodo, bacI-A1, PnecC, betIII-A1, acI-B2, and acI-B4. These results show that despite lake-to-lake differences and interannual variability, there are bacterial taxa that are consistently present in bog lakes. We note that tribes correspond very roughly to species-level designations as explained previously.

Principal coordinates analysis suggested that samples clustered also by mixing regime. We thus evaluated Venn diagrams of OTUs shared by, and unique to, each mixing regime to better visualize the overlap in community composition (Figure 3.7). In both epilimnia and hypolimnia, meromictic lakes had the greatest numbers of unique OTUs while polymictic lakes had the least, consistent with the differences in richness between lakes. Shared community membership, i.e. the number of OTUs present at any abundance in both communities, differed between mixing regimes. Epilimnia (A) and hypolimnia (B) showed similar trends in shared membership: meromictic and dimictic lakes shared the most OTUs, while meromictic and polymictic lakes shared the least.

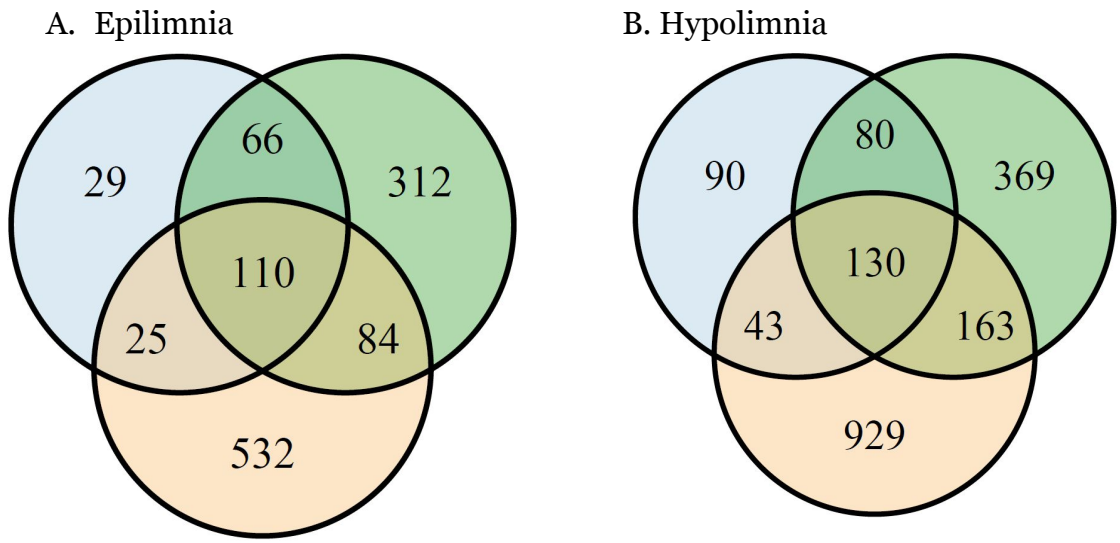

Figure 3.7. Numbers of unique and shared OTUs by mixing regime. To better understand how shared community membership differs by mixing regime, we quantified the number of shared and unique OTUs in each category. An OTU needed only to appear in one sample at any abundance to be considered present in a category. We found that in both layers, meromictic lakes have the greatest numbers of unique OTUs and polymictic lakes have the least. Meromictic and dimictic lakes shared the most OTUs, while meromictic and polymictic lakes shared the least.

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

Actinobacteria). An abundance threshold of 500 reads was imposed on each taxonomic group.

The lineage acI is a ubiquitous freshwater group, with specific clades and tribes showing a preference for bog lakes in previous studies (92, 108). Our dataset shows a further distinction of acI by mixing regime in epilimnia; acI-A tribes were found predominantly in meromictic lakes, with exception of Phila, which is an indicator of polymictic lakes. Tribes of acI-B, particularly OTUs belonging to acI-B2, were indicators of dimictic lakes. *Methylophilales*, a putative methylotroph (109), was also an indicator of dimictic lakes, as was the putative sulfate reducing family *Desulfobulbaceae*. The phyla *Planctomyces, Omnitrophica* (formerly OP3), OP8, and *Verrucomicrobia* were found more often in meromictic lakes, as were putative sulfate reducing taxa belonging to *Syntrophobacterales* and *Desulfobacteraceae*. Indicators of polymictic lakes include ubiquitous freshwater groups such as *Limnohabitans*, *Polynucleobacter* (PnecC), betI-A, and verI-A. Thus, despite the observed variability and differences between lakes, layers, and years, we detected a core community composed of ubiquitous freshwater bacterial groups and identified indicator taxa endemic to groups of sites defined by mixing frequency.

Lifestyles of freshwater lineages

Because of the observed variability in bacterial community dynamics, we next asked if individual OTUs showed consistent levels of abundance, persistence, and variability. We defined these metrics as mean abundance when present, the proportion

of samples containing the group of interest, and the coefficient of variation for lineages classified using the freshwater taxonomy, respectively. These metrics have been previously used to categorize OTUs (33, 110). Using only well-defined freshwater groups allowed better taxonomic resolution as we summed the abundances of OTUs by their lineage classification. We note that lineage is very roughly analogous to family in our provisional freshwater taxonomy. Lifestyle traits of lineages were consistent across both lakes and years. Low persistence was associated with high variability, and low variability was associated with high abundance (Figure 3.8). We rarely observed "bloomers," situations where a clade had both high abundance and low persistence; one potential reason for this could be that true "bloomers" drop below the detection limit of our sequencing methods when not abundant. Most freshwater lineages were highly persistent at low abundances with low variability. Lineage gamIII of the Gammaproteobacteria was an exception, with low persistence, low abundance, and high variability. Lineages gamI and verI-A occasionally also exhibited this profile. Lineages betII and acI were highly abundant and persistent with low variability, consistent with their suggested lifestyles as ubiquitous freshwater generalists (36, 92). Even though OTUs did not show the same abundance dynamics each year, they did exhibit patterns that are consistent between years and lakes.

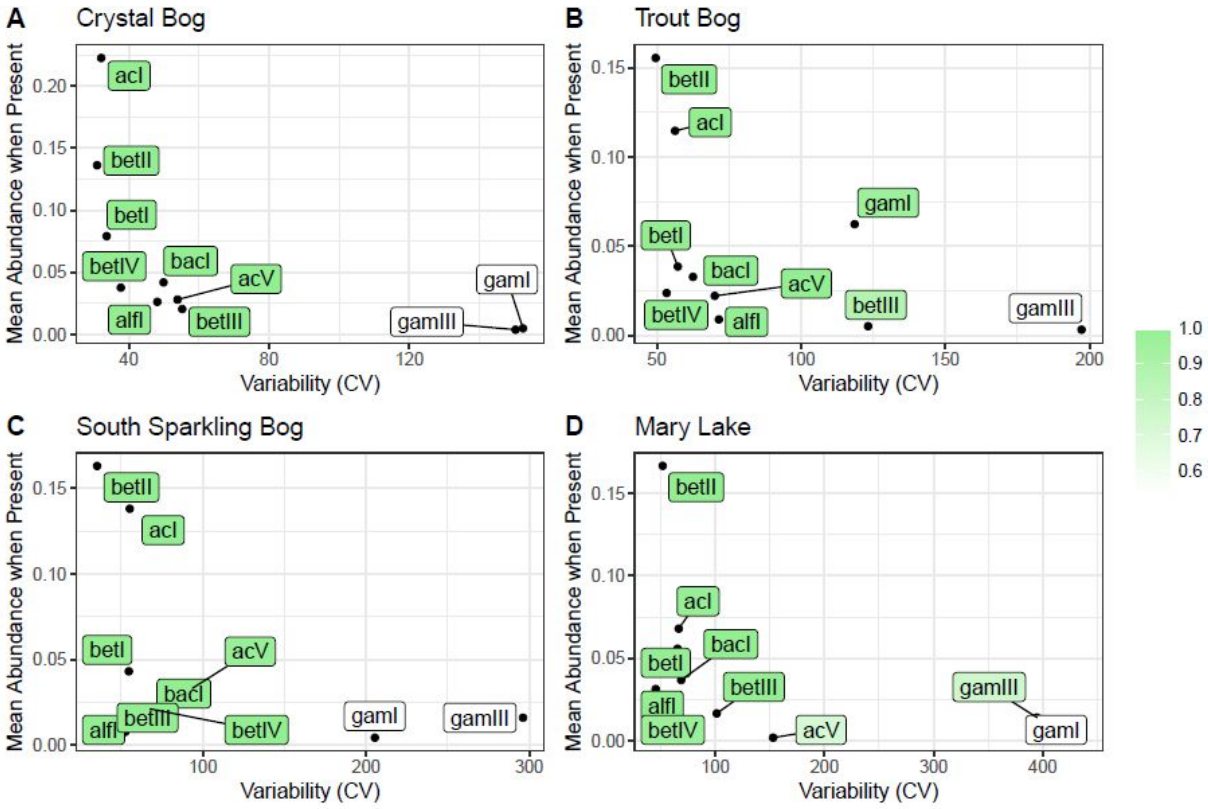

Figure 3.8. Traits of freshwater lineages. These well-defined freshwater groups showed similar persistence, variance, and abundance in every lake, despite differing abundance patterns. Data from epilimnia with at least two years of undisturbed sampling are shown here. Mean abundance was represented as the average percentage of reads attributed to each lineage when that lineage was present. Variability was measured as the coefficient of variation. Persistence (shaded color) was defined as the proportion of samples containing each lineage.

Discussion

The North Temperate Lakes - Microbial Observatory bog dataset is a comprehensive 16S rRNA gene amplicon survey spanning four years, eight lakes, and two thermal layers. We hypothesized that alpha and beta diversity would be associated with mixing frequency in bog lakes. Richness and membership in these communities

were structured by layer, mixing regime, and lake. However, we found that multiple years of sampling were necessary to census the community of bog lake ecosystems. We identified specific bacterial taxa core to bog lakes, as well as taxa endemic to certain depths or mixing regimes. High levels of variability were detected in this dataset; the community composition observed in each lake and each year of sampling was unique. However, freshwater lineages still showed consistent lifestyles, defined by abundance, persistence, and variability, across lakes and years, even though the abundance trends of individual OTUs varied each year. Our results emphasize the importance of multiple sampling events to assess full bacterial community membership and variability in an ecosystem.

We supported previous research on the characteristics of bacterial communities in the epilimnion and hypolimnion and the association of lake mixing frequency with community composition. We confirmed that epilimnia communities tended to be more dispersed than hypolimnia communities, potentially due to increased exposure to climatic events (95). Mixing was disruptive to both epilimnion and hypolimnion communities, selecting for only a few taxa that persist during this disturbance, but quickly recovering diversity once stratification was re-established (44). Our initial inspiration for the collection of this dataset was the intermediate disturbance hypothesis. We hypothesized that water column mixing is a disturbance to bog lake bacterial communities, and that lakes with intermediate mixing frequency would have the highest levels of biodiversity. Comparing richness between lakes of different mixing

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

We were not able to detect repeatable annual trends in bog lakes in our multiple years of sampling. While seasonality in marine and river systems has been well-established by our colleagues, previous research on seasonality in freshwater lakes has produced inconsistent results (6, 111–113). Distinct, seasonally repeatable community types were identified in alpine lakes, but stratified summer communities were distinct each year (114). Seasonal trends were detected in a time series from Lake Mendota similar to this study, but summer samples in Lake Mendota were more variable then those collected in other seasons (89). In the previous ARISA-based research on the bog lakes in our dataset, community properties such as richness and rate of change were consistent each year, and the phytoplankton communities were hypothesized to drive seasonal trends in the bacterial communities based on correlation studies (56, 57, 76). Synchrony in seasonal trends was observed; however, in a second year of sampling for seasonal trends in Crystal Bog and Trout Bog, these findings were

not reproduced (115). Successional trends were studied in Crystal Bog and Lake Mendota with a relatively small number of samples collected over two years and "dramatic changes" in community composition associated with drops in biodiversity were described during the summer months, while spring, winter, and fall had more stable community composition (56). Because our dataset was sparsely represented by seasons other than summer, higher summer variability may explain why we see a different community each year and a lack of seasonal trends in community composition. However, we cannot disprove the influence of seasonality on bacterial community dynamics in temperate freshwater lakes as a general feature. Our results may indeed point to a feature that is unique to darkly stained acidic bog lakes. Even in marine systems, trends in seasonality differ by site and OTU definition, and continued long term time series sampling is suggested as an approach needed to elucidate these trends and link seasonality in bacterial community composition to biogeochemical cycling.

One of the biggest benefits of 16S rRNA gene amplicon sequencing over ARISA is the ability to assign taxonomic classifications to sequences. Tracking bacterial taxa through multiple sites and over multiple years allowed us to detect consistent lifestyle trends, despite a lack of predictability in seasonal trends. Some groups, such as acI (*Actinobacteria*) and betII (*Betaproteobacteria*), were persistent, abundant, and not variable, much like *Pelagibacter ubique* (SAR11) in marine systems. Other freshwater taxa such as gamI and gamIII (both *Gammaproteobacteria*) exhibited a pattern of low persistence, low abundance, and high variability. Unlike in the oceans, where taxa such

as *Alteromonas* "bloom and bust" (116), no taxa classified within the freshwater taxonomy with high abundance and low persistence or high variability were observed. This suggest that either bog lakes are not conducive to the large blooms of a single population as observed in other freshwater lakes, or that taxa with this lifestyle dropped below our detection limit when not blooming.

In addition to a core of persistent taxa found in nearly every sample collected, we also identified taxa endemic to either the epilimnion or hypolimnion and to specific mixing regimes. These endemic taxa likely reflect the physical and/or biogeochemical differences driven by mixing regime. Dimictic and meromictic hypolimnia, which are consistently anaerobic, harbor putative sulfur cycling groups not present in polymictic hypolimnia, which are more frequently oxygenated. Members of the acI lineage partition by mixing regime in epilimnia, and the functional traits driving this filtering effect are the subject of active study (117). Interestingly, the meromictic Mary Lake hypolimnion contains several taxa classified into the candidate phyla radiation (118) and a larger proportion of completely unclassified reads than other hypolimnia. This is consistent with the findings of other 16S rRNA gene amplicon sequencing and metagenomics studies of meromictic lakes (119, 120), and suggests that the highly reduced and consistently anaerobic conditions in meromictic hypolimnia are excellent study systems for research on members of the candidate phyla radiation and "microbial dark matter."

Perhaps the biggest implication of this study is the importance of repeated sampling of microbial ecosystems. A similar dataset spanning only a single year would not have captured the full extent of variability observed, and therefore would not have detected as many of the taxa belonging to the bog lake community; even our four years of weekly sampling during the summer stratified period did not result in level rarefaction curves. While we found no evidence for seasonal trends or repeated annual trends, it is possible that there are cycles or variables acting on scales longer than the five years covered in this dataset, or that interannual differences are driven by environmental factors that do not occur every year. Unmeasured biotic interactions between bacterial taxa may also contribute to the observed variability. Understanding the factors that contribute to variability in lake communities will lead to improved predictive modelling in freshwater systems, allowing forecasting of bloom events and guiding better management strategies. Additionally, these 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 continue to collect and sequence samples for the North Temperate Lakes – Microbial Observatory, and we are expanding our sequencing repertoire beyond 16S rRNA gene amplicon sequencing. All 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. This dataset

has already been used in a meta-analysis of microbial time series (121) and a synthesis of microbial community composition across a multitude of ecosystems (122). 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.

Chapter 4: Connections between freshwater carbon and nutrient cycles revealed through reconstructed population genomes

Alexandra M. Linz[1], Shaomei He[1,2], Sarah L. R. Stevens[1], Karthik Anantharaman[1], Robin R. Rohwer[3], Rex R. Malmstrom[4], Stefan Bertilsson[5], Katherine D. McMahon[1,6]

[1]Department of Bacteriology, University of Wisconsin–Madison, [2]Department of Geoscience, University of Wisconsin-Madison, [3]University of Wisconsin-Madison Environmental Chemistry and Technology Program, [4]Department of Energy Joint Genome Institute, [5]Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University, [6]Department of Civil and Environmental Engineering, University of Wisconsin–Madison

A.M. Linz designed experiments, analyzed data, and wrote manuscript
S. He contributed analysis tools, advised analyses, and edited manuscript
S.L.R. Stevens analyzed data, advised analyses, and edited manuscript
K. Anantharaman contributed analysis tools, advised analyses, and edited manuscript
R.R. Rohwer analyzed data and edited manuscript
R.R. Malmstrom designed experiments, advised analyses, and edited manuscript
S. Bertilsson designed experiments, advised analyses, and edited manuscript
K.D. McMahon designed experiments, advised analyses, and wrote manuscript

Publication: Linz, A., He, S., Stevens, S. L., Anantharaman, K., Rohwer, R. R., Malmstrom, R., Bertilsson, S., & McMahon, K. D. (2018). Connections between freshwater carbon and nutrient cycles revealed through reconstructed population genomes. *bioRxiv*, 365627. Submitted to Peer J on July 3rd, 2018.

Introduction

Lakes collect nutrients from surrounding terrestrial ecosystems (123), placing lakes as "hotspots" in the landscape for carbon nutrient cycling (124). Much of this biogeochemical cycling is performed by freshwater microbes. We have learned much about freshwater microbes through previous research that has revealed the geographic distribution of freshwater taxa (125), the distribution of functional marker genes (15, 61, 62, 85), and substrate use capabilities in specific phylogenetic groups (126). However, organism level information about microbial metabolisms is currently not well incorporated into conceptual models of freshwater nutrient cycling.

Previously, we used time series metagenomics to assemble nearly 200 metagenome-assembled genomes (MAGs) from two temperate lakes: Lake Mendota, a highly productive eutrophic lake, and Trout Bog, a humic bog lake. These MAGs were used to study genome-wide diversity sweeps in Trout Bog (54), to build metabolic networks of the ubiquitous freshwater *Actinobacteria* acI in Lake Mendota (117), and to propose functions for freshwater *Verrucomicrobia* (35). In addition to this body of knowledge based on the MAG dataset, previous time series of analyses of 16S rRNA gene amplicon datasets from both lakes provide an understanding of taxon dynamics over time (45, 90). Lake Mendota and Trout Bog are ideal sites for comparative time series metagenomics because of their history of extensive environmental sampling by the North Temperate Lakes - Long Term Ecological Research program (NTL-LTER, http://lter.limnology.wisc.edu), and their contrasting limnological attributes (Table

4.1). Here, we build on this previous work by identifying contrasting patterns of nutrient cycling between lakes based on analyses of functional marker genes and MAGs.

Gene-centric methods are one approach that can identify community functions, while analysis of population genomes using MAGs can identify coupled metabolic processes taking place within the boundary of a cell. In this research, we use functional marker genes and MAGs from two freshwater lakes with contrasting chemistry to yield insights about microbial metabolisms in freshwater ecosystems. We identified genes and pathways purportedly involved in primary production, DOC mineralization, and nitrogen and sulfur cycling. Some types of metabolism were found in both sites despite their different chemistry profiles, but in different taxonomic groups. We demonstrate how MAGs and metagenomic time series can be used to track specific phylogenetic groups capable of key biogeochemical transformations. Finally, we introduce the MAG collection as a valuable community resource for other freshwater microbial ecologists to mine and incorporate into comparative studies across lakes around the world.

Table 4.1. Characteristics of Lake Mendota and Trout Bog. Water from Lake Mendota and Trout Bog was sampled weekly during the ice-free periods using an integrated water column sampler and filtered for DNA using a 0.22 micron filter. Metagenomic sequencing was performed on DNA extracted from filters collected in 2008-2012 from Lake Mendota and in 2007-2009 from Trout Bog. The epilimnion (upper thermal layer) was sampled in both lakes, while the hypolimnion (bottom thermal layer) was sampled only in Trout Bog. Chemistry data were collected by NTL-LTER from depth discrete samples taken from 0 and 4 m for Lake Mendota, 0 m for the Trout Bog Epilimnion, and 3 and 7 m for the Trout Bog Hypolimnion. Values reported here are the means of all measurements in the sampling time span for each lake, with standard deviations reported in parentheses.

	Lake Mendota	Trout Bog Epilimnion	Trout Bog Hypolimnion
Location	Madison, WI	Boulder Junction, WI	
Coordinates	43.107055, -89.411729	46.041172, -89.686297	
Depth of lake (m)	25.3	7.9	
Surface area of lake (km^2)	39.61	0.01	
Microbial sampling depth range (m)	0-12	0-2	2-7
Years sampled	2008-2012	2007-2009	2007-2009
Oxygenation	Oxic	Oxic	Suboxic/Anoxic
pH	8.6 (0.4)	5.0 (0.2)	5.3 (0.2)
Dissolved inorganic carbon (ppm)	40.8 (5.0)	2.6 (2.2)	6.9 (3.1)
Dissolved organic carbon (ppm)	6.0 (6.2)	17.7 (5.2)	22 (6.2)
Total dissolved nitrogen (ppb)	923 (487)	637 (204)	1392 (1031)
Total nitrogen (ppb)	1098.81 (520.92)	831.49 (315.65)	1684 (1563)
Total dissolved phosphorus (ppb)	43.92 (50.97)	15.00 (13.65)	69 (98)
Total phosphorus (ppb)	64.11 (52.30)	32.29 (13.65)	95 (126)
Sulfate (ppm)	16.9 (1.5)	1.2 (0.3)	0.9 (0.7)

Methods

Sample collection

Samples were collected from Lake Mendota and Trout Bog as previously described (54). Briefly, integrated samples of the water column were collected during the ice-free periods of 2007-2009 in Trout Bog and 2008-2012 in Lake Mendota. In Lake Mendota, the top 12 meters of the water column were sampled, approximating the epilimnion (upper, oxygenated, and warm thermal layer). The epilimnion and hypolimnion (bottom, anoxic, and cold thermal layer) of Trout Bog were sampled separately at depths determined by measuring temperature and dissolved oxygen concentrations throughout the water column; the sampling depths were most often 0-2 meters for the epilimnion and 2-7 meters for the hypolimnion. DNA was collected by filtering 150 mL of the integrated samples on 0.2-um pore size polyethersulfone Supor filters (Pall Corp., Port Washington, NY, USA). Filters were stored at -80C until extraction using the FastDNA Kit (MP Biomedicals, Burlingame, CA, USA).

Sequencing

As previously described (54), metagenomic sequencing was performed by the Department of Energy Joint Genome Institute (DOE JGI) (Walnut Creek, CA, USA). A total of 94 samples were sequenced for Lake Mendota, while 47 metagenomes were sequenced for each layer in Trout Bog. Samples were sequenced on the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA), except for four libraries (two from each

layer of Trout Bog) sequenced using the Illumina TruSeq protocol on the Illumina GAIIx platform (Illumina). Paired-end sequencing reads were merged with FLASH v1.0.3 with a mismatch value of less than 0.25 and a minimum of 10 overlapping bases, resulting in merged read lengths of 150-290 bp (127). 16S rRNA amplicon iTag sequencing was also performed on these samples. This data is available under DOE JGI project IDs 1078703 and 1018581 for Trout Bog and Lake Mendota, respectively. Samples from Trout Bog were sequenced on the 454 GS FLX-Titanium platform (Roche 454, Branford, CT, USA); the V8 hypervariable region (primer 1392R: acgggcggtgtgtRc) (128) and sequences were trimmed to 324 base pairs using VSEARCH (v2.3.4) (129). Samples from Lake Mendota were sequenced on an Illumina MiSeq, and the V4 region was targeted using paired-end sequencing (primers 525F: GTGCCAGCMGCCGCGGTAA and 806R: GGACTACHVGGGTWTCTAAT) (97). Both datasets were trimmed based on alignment quality and chimera checking using mothur v.1.39.5 (99). Unclustered unique sequences were assigned taxonomy using TaxAss (130) to leverage the FreshTrain (32) and Greengenes (version 13_5) (100).

Assembly and Binning

To recover MAGs, metagenomic reads were pooled by lake and layer and then assembled as previously described (Table 4.2) (54, 131). In Trout Bog, this assembly was performed using SOAPdenovo2 at various k-mer sizes (132) and the resulting contigs were combined using Minimus (133). In Lake Mendota, merged reads were assembled using Ray v2.2.0 with a single k-mer size (134). Contigs from the combined assemblies

were binned using MetaBAT (-veryspecific settings, minimum bin size of 20kb, and minimum contig size of 2.5kb) (135) and reads from unpooled metagenomes were mapped to the assembled contigs using the Burrows-Wheeler Aligner (≥ 95% sequence identity, n = 0.05) (136), which allowed time-series resolved binning. DOE JGI's Integrated Microbial Genome (IMG) database tool (https://img.jgi.doe.gov/mer/) (137) was used for gene prediction and annotation. Annotated MAGs can be retrieved directly from the IMG database and JGI's Genome Portal using the IMG Genome ID provided (also known as IMG Taxon ID). MAG completeness and contamination/redundancy was estimated based on the presence of a core set of genes with CheckM (70, 138), and MAGs were classified using Phylosift (139).

Functional Marker Gene Analysis

To analyze functional marker genes in the unassembled, unpooled metagenomes, we used a curated database of reference protein sequences (140) and identified open reading frames (ORFs) in our unassembled metagenomic time series using Prodigal (141). This analysis was conducted on merged reads. The protein sequences and ORFs were compared using BLASTx (142) with a cutoff of 30 percent identity. Significant differences in gene frequency between sites were identified using LEfSE (143). Read abundance was normalized by metagenome size for plotting. We chose to perform this analysis because gene content in unassembled metagenomes is likely more quantitative and more representative of the entire microbial community than gene content in the MAGs.

Table 4.2. Statistics from genome assembly and binning. Metagenomic samples were pooled by lake and layer to allow time-resolved binning. The time series in Lake Mendota spans 2008-2012, while the Trout Bog time series spans 2007-2009. The large amount of DNA assembled produced just under 200 high quality metagenome-assembled genomes.

	Lake Mendota	Trout Bog Epilimnion	Trout Bog Hypolimnion
Number of metagenomes	94	47	47
Collection time span	Jun. 2008 – Nov. 2012	Jun. 2007 – Aug. 2009	May 2007 – Aug. 2009
Total base pairs in metagenomes	1.26×10^{11}	6.72×10^{10}	7.18×10^{10}
Total base pairs in pooled assembly	3.37×10^{9}	2.60×10^{8}	5.47×10^{8}
Number of contigs in pooled assembly	9,912,431	79,862	153,912
Number of curated bins	99	31	63
Number of base pairs in curated bins	2.31×10^{8}	5.82×10^{7}	1.60×10^{8}
Number of contigs in curated bins	18,675	5,098	11,656

Pathway Prediction

Only MAGs that were at least 50% complete with less than 10% estimated contamination (meeting the MIMARKS definition of a medium or high quality MAG) were included in this study (144). Pathways were analyzed by exporting IMG's functional annotations for the MAGs, including KEGG, COG, PFAM, and TIGRFAM annotations and mapped to pathways in the KEGG and MetaCyc databases as previously described (35). To score presence, a pathway needed at least 50% of the required enzymes encoded by genes in a MAG and if there were steps unique to a pathway, at least one gene encoding each unique step. Putative pathway presences was aggregated by lake and phylum in order to link potential functions identified in the metagenomes to taxonomic groups that may perform those functions in each lake. Glycoside hydrolases were annotated using dbCAN (http://csbl.bmb.uga.edu/dbCAN) (145). Nitrogen usage in amino acids was calculated by taking the average number of nitrogen atoms in translated ORF sequences across each MAG.

Data formatting and plotting was performed in R (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.) using the following packages: ggplot2 (H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.), cowplot (Claus O. Wilke (2017). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 0.9.2. https://CRAN.R-project.org/package=cowplot), reshape2 (Hadley Wickham (2007).

Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL http://www.jstatsoft.org/v21/i12/.), and APE (Paradis E., Claude J. & Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289-290.). The datasets, scripts, and intermediate files used to predict pathway presence and absence are available at <https://github.com/McMahonLab/MAGstravaganza>. Any future updates or refinements to this dataset will be available at this link.

Results/Discussion

Community Functional Marker Gene Analysis

To assess potential differences in microbial metabolisms between Lake Mendota and Trout Bog, we tested whether functional marker genes identified in the unassembled merged metagenomic reads appeared more frequently in one lake or layer compared to the others (Figure 4.1, Table 4.3). These comparisons were run between the epilimnia of Trout Bog and Lake Mendota, and between the epilimnion and hypolimnion of Trout Bog. We did not compare the epilimnion of Lake Mendota to the hypolimnion of Trout Bog, as the multitude of factors differing between these two sites make this comparison illogical. Many genes differed significantly by site, indicating contrasting gene content between lakes and layers. To further infer differences in microbial metabolism, we aggregated marker genes by function (as several marker genes from a phylogenetic range were included in the database for each type of function)

and tested for significant differences in distribution between lakes and layers using a Wilcoxon rank sum test with a Bonferroni correction for multiple pairwise testing. Many functional markers were found to be significantly more abundant in specific sites; more will be reported in each of the following sections. These contrasting abundances of functional marker genes suggest significant differences in the metabolisms of microbial communities across lake environments.

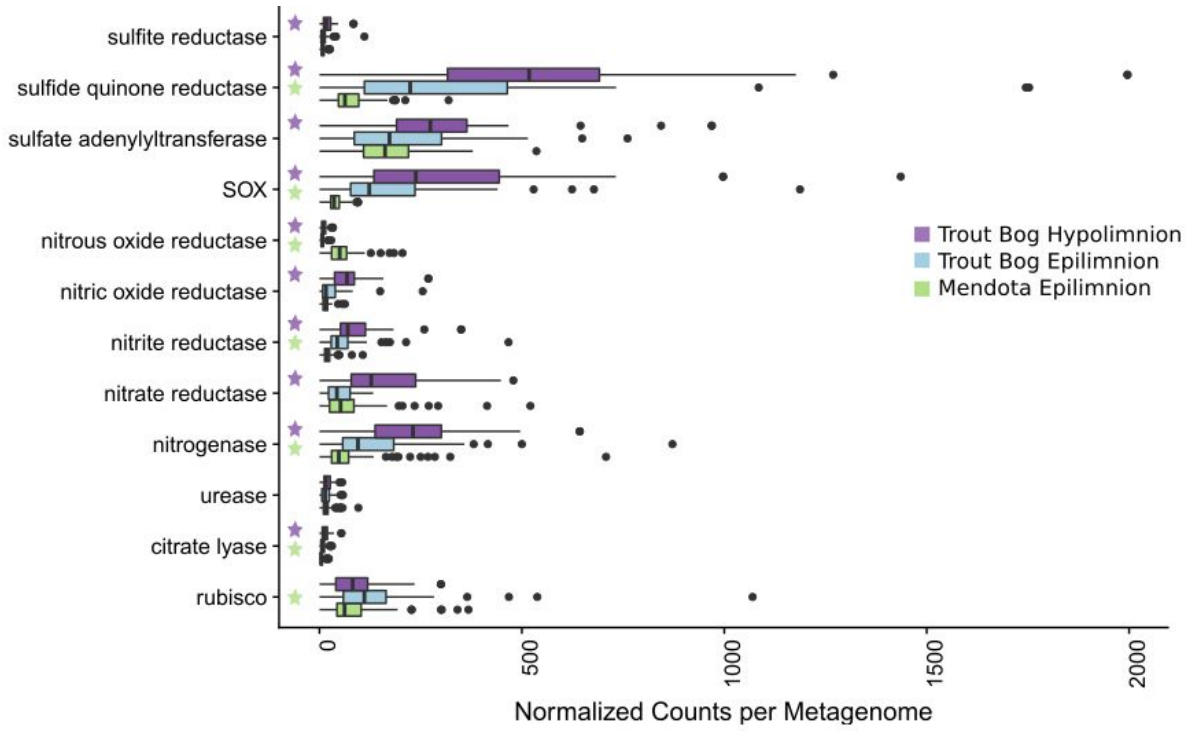

Figure 4.1. Analysis of marker gene abundances reveals differences between lakes and layers. To assess potential differences in microbial metabolisms in our study sites, we predicted open reading frames in unassembled metagenomes using Prodigal and compared the resulting ORFs to a custom database of metabolic marker genes using BLAST. In these boxplots, significant differences in numbers of gene hits between sites was tested using a pairwise Wilcoxon rank sum test with a Bonferroni correction; significance was considered to be $p < 0.05$. 94 metagenomes were tested for Lake Mendota, while 47 metagenomes were tested in each layer of Trout Bog. Significant differences between the Trout Bog and Lake Mendota epilimnia and between the Trout Bog epilimnion and hypolimnion are indicated by a green or a purple star, respectively. Significant differences between the Trout Bog hypolimnion and the Lake Mendota epilimnion were not tested, as the large number of variables differing in these sites makes the comparison less informative. This analysis revealed differences in the number of marker genes observed by lake for many metabolic processes involved in carbon, nitrogen, and sulfur cycling. P-values of markers described in Figure 4.1 and elsewhere in the text are reported in Table 4.3.

Table 4.3. P-values of marker gene distributions between sites. A Wilcoxon rank sum test was used to non-parametrically test for significant differences in functional marker gene distributions between our study sites. P-values of less than 0.05 are considered significant.

Functional Marker	Mendota vs Trout Bog Epilimnion	Trout Bog Epilimnion vs Hypolimnion
RubisCO	0	0
Citrate lyase	0.01	0.01
Urease	0.69	1
Nitrogenase	0	0
Nitrate reductase	1	0
Nitrite reductase	0	0
Nitric oxide reductase	0.83	0
Nitrous oxide reductase	0	0.01
SOX	0	0.04
Sulfate adenylyltransferase	0.94	0
Sulfide quinone reductase	0	0
Sulfite reductase	1	0

How Representative are the MAGs?

To identify the phylogenies of the microbes carrying marker genes and the co-occurrences of marker genes within the same population genomes, we used metagenome-assembled genomes (MAGs) from each metagenomic time series to predict metabolic pathways based on genomic content. A total of 193 medium to high quality bacterial MAGs were recovered from the three combined time series metagenomes in Trout Bog and Lake Mendota: 99 from Lake Mendota, 31 from Trout Bog's epilimnion, and 63 from Trout Bog's hypolimnion. These population genomes ranged in estimated completeness from 50 to 99% based on CheckM estimates (146). Several MAGs from Trout Bog's epilimnion and hypolimnion appeared to belong to the same population based on average nucleotide identities greater than 99% calculated using JGI's ANI calculator (147). This is likely because assembly and binning were carried out separately for each thermal layer, even though some populations were present throughout the water column. To assess the diversity of our MAGs, we constructed an approximate maximum likelihood tree of all the MAGs in FastTree (148) using whole genome alignments (Figure 4.2). The tree is not intended to infer detailed evolutionary history, but to provide an overall picture of similarity between genomes. MAGs recovered are a diverse set of genomes assigned to taxa typically observed in freshwater.

The phylum-level assignments of our MAGs were largely matched the classifications of 16S rRNA gene amplicon sequencing results averaged across the time series, consistent with a higher likelihood of recovering MAGs from the most abundant

populations in the community (Figure 4.3). However, some taxa, including *Tenericutes, Ignavibacteria, Epsilonproteobacteria*, and *Chlamydiae*, were represented by MAGs but not identified in the 16S gene amplicon datasets. *Chlorobi* was overrepresented by MAG coverage compared to 16S rRNA gene counts, while *Proteobacteria* was overrepresented by 16S rRNA gene counts compared to MAG coverage. These discrepancies could be explained by bias in the 16S primer sets (149) difference in rRNA copy number, or assembly bias in MAG recovery. The observed taxonomic compositions are consistent with other 16S-based studies from these lakes (45, 90). The detection of similar phyla using both methods suggests that our MAGs are representative of the resident microbial communities.

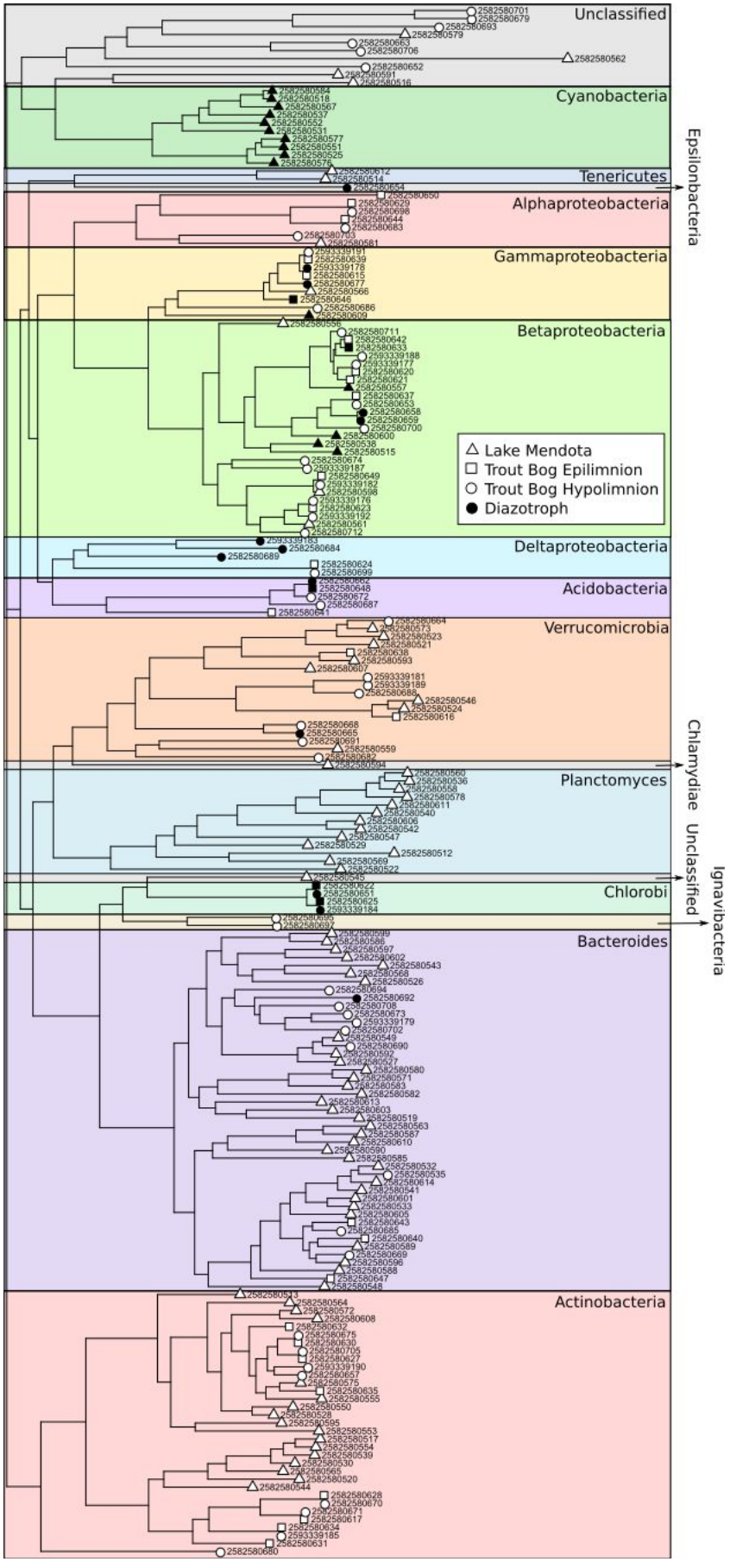
Unclassified
Cyanobacteria
Tenericutes
Alphaproteobacteria
Gammaproteobacteria
Betaproteobacteria
Deltaproteobacteria
Acidobacteria
Verrucomicrobia
Planctomyces
Chlorobi
Bacteroides
Actinobacteria
Epsilonbacteria
Chlamydiae
Unclassified
Ignavibacteria
Lake Mendota
Trout Bog Epilimnion
Trout Bog Hypolimnion
Diazotroph

Figure 4.2. Tree of diversity and nitrogen fixation in our MAGs. To visualize the diversity of our MAGs, phylogenetic marker genes were extracted from each MAG and aligned using Phylosift. An approximate maximum-likelihood tree based on these alignments was constructed using FastTree. The potential for nitrogen fixation based on gene content is indicated on the branch tips.

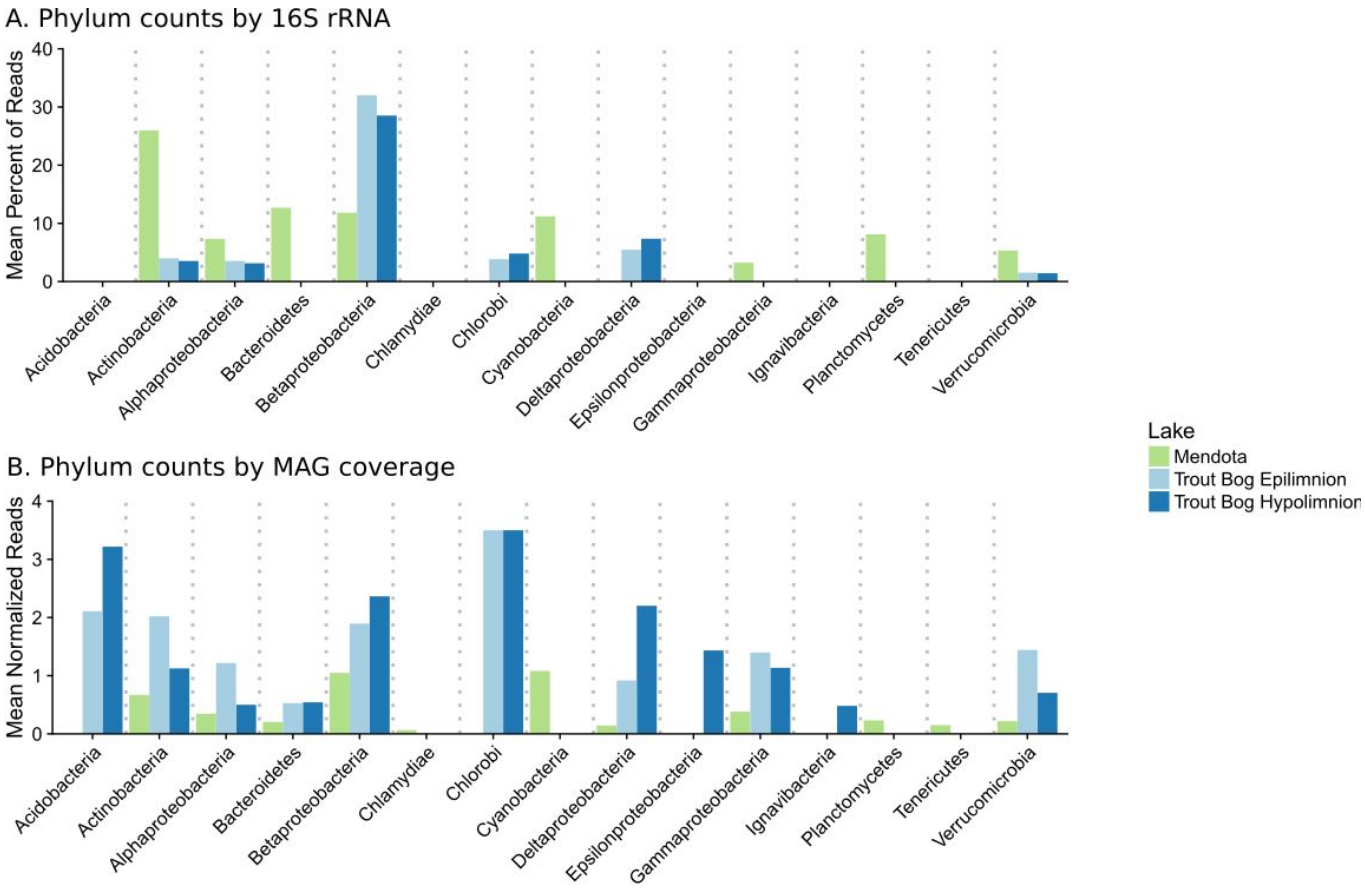

Figure 4.3. How representative are the MAGs of the microbial communities? The community composition observed via 16S rRNA gene amplicon sequencing (A) and inferred using the proportions of reads from the same metagenomic time series samples that mapped to set of MAGs affiliated with major phyla (B). MAGs were classified using Phylosift, while 16S sequences were classified to the phylum level. Numbers above bars indicating abundances greater than the limit of the y-axis. The 16S V6-V8 region was targeted in Trout Bog, while the V4 region was targeted in Lake Mendota. *Proteobacteria* was split into classes due to the high diversity of this phylum. Although proportions vary, similar taxonomic groups are observed using both approaches. Differences are likely due to a combination of primer and assembly biases. However, similar phyla were detected using both methods, suggesting that our MAG datasets are representative of their communities.

Nitrogen Cycling

Nitrogen availability is an important factor structuring freshwater microbial communities. To see if there were differences in nitrogen cycling between different lake environments, we analyzed nitrogen-related marker genes and the MAGs containing nitrogen cycling pathways. We discovered significant differences in the abundances of marker genes (Figure 4.1), along with phylogenetic difference in the populations containing these pathways.

To identify differences in nitrogen fixation between sites, we analyzed marker genes encoding nitrogenase subunits. Genes encoding for nitrogenase were observed most frequently in metagenomes from Trout Bog's hypolimnion, followed by the Trout Bog's epilimnion, and lastly by Lake Mendota's epilimnion. The nitrogenase enzyme is inhibited by oxygen, which could explain the higher abundance of nitrogenase anoxic hypolimnion of Trout Bog. We further analyzed MAGs containing genes encoding nitrogenase and found differences in the taxonomy of putative diazotrophs between the two ecosystems (Figure 4.2). In Lake Mendota, two thirds of MAGs encoding the nitrogen fixation pathway were classified as *Cyanobacteria,* while the other third was assigned to *Betaproteobacteria* and *Gammaproteobacteria*. Although not all *Cyanobacteria* fix nitrogen, previous measurements of nitrogen fixation in Lake Mendota found a strong correlation between this pathway and the *Cyanobacteria Aphanizomenon* (150). MAGs containing genes encoding nitrogen fixation were more phylogenetically diverse in Trout Bog and included *Deltaproteobacteria,*

Gammaproteobacteria, Epsilonproteobacteria, Acidobacteria, Verrucomicrobia, Chlorobi, and *Bacteroidetes*. The increased diversity of diazotrophs in Trout Bog compared to Lake Mendota suggests that nitrogen fixation genes may be horizontally transferred with populations in Trout Bog.

To identify differences in denitrification, we analyzed marker genes for denitrification, including reductases for nitrous oxide, nitric oxide, nitrite, and nitrate. These denitrification genes had a similar trend as the nitrogen fixation genes; they were observed most frequently in metagenomes from the Trout Bog hypolimnion, with the exception of nitrous oxide reductase, which was most frequently found in Lake Mendota. This trend could stem from denitrification also requiring a reductive, low oxygen environment. Urease, another nitrogen cycling marker gene, was not found significantly more often in any site. We further analyzed putative denitrification pathways in our MAGs and found that they were observed at similar frequencies in population genomes from all environments (Figure 4.4). Urea degradation pathways were also predicted in MAGs from both lakes, which is consistent with research showing that urea is a common nitrogen source for bacteria in multiple freshwater environments (151–153)

To explore the importance of polyamines in the freshwater nitrogen cycle, we analyzed genes encoding the biosynthesis and degradation of polyamines such as spermidine and putrescine. We predicted that 94% of MAGs could synthesize polyamines, and 87% could degrade polyamines. These genes were prevalent in many

diverse MAGs from both lakes, including *Actinobacteria* as has been previously observed (92, 117). While there is some evidence for the importance of polyamines in aquatic systems (154), the ecological role of these compounds in freshwater is not fully resolved. Polyamines are known to play a critical but poorly understood role in bacterial metabolism (155), and the exchange of these nitrogen compounds between populations may be a factor structuring freshwater microbial communities. Polyamines can also result from the decomposition of amino acids, so higher trophic levels such as fish or zooplankton may provide an additional source (156). The frequent appearance of polyamine-related pathways in our MAGs lends support to the hypothesis that these compounds are important parts of the dissolved organic nitrogen and carbon pool in freshwater.

To identify signatures of nitrogen limitation at the genomic level, we analyzed biases in amino acid use in our MAGs (157, 158). For this analysis, genomes from the Trout Bog layers were considered together due to the previously mentioned overlap in recovered genomes. We observed on average, MAGs from Trout Bog encoded amino acids with 1% less nitrogen than MAGs from Lake Mendota. Although this difference is small, it was significant using a Wilcoxon rank sum test ($p = 0.02$). The observed amino acid bias suggests that conditions in Trout Bog may lead to stronger selection for nitrogen poor proteins than in Lake Mendota. Differences in the compositions of the nitrogen pools in these lakes may also contribute to the observed differences in the distributions of nitrogen cycling marker genes. Lake Mendota receives large amounts of

nitrate runoff from the surrounding agricultural landscape, while Trout Bog receives nitrogen in more complex forms (e.g. *Sphagnum*-derived organic nitrogen), and the microbial community competes for nitrogen with the surrounding plant community.

Sulfur Cycling

Sulfur is another essential element in freshwater that is cycled between oxidized and reduced forms by microbes. Our marker gene analysis demonstrated that genes encoding for sulfide:quinone reductase (for sulfide oxidation) and the sox pathway (for thiosulfate oxidation) were significantly more abundant in Trout Bog compared to Lake Mendota, with no significant differences between the layers of Trout Bog (Figure 4.1). Genes encoding for sulfite reductases were the least abundant sulfur cycling marker genes in all sites. Dissimilatory sulfite reductase was observed only in MAGs from Trout Bog, especially those classified as *Chlorobiales*. Because this enzyme is thought to operate in reverse in green sulfur-oxidizing phototrophs such as *Chlorobiales* (159), this may indicate an oxidation process rather than a reductive sulfur pathway. Assimilatory sulfate reduction was the most common sulfur-related pathway identified in the MAGs (Figure 4.4).

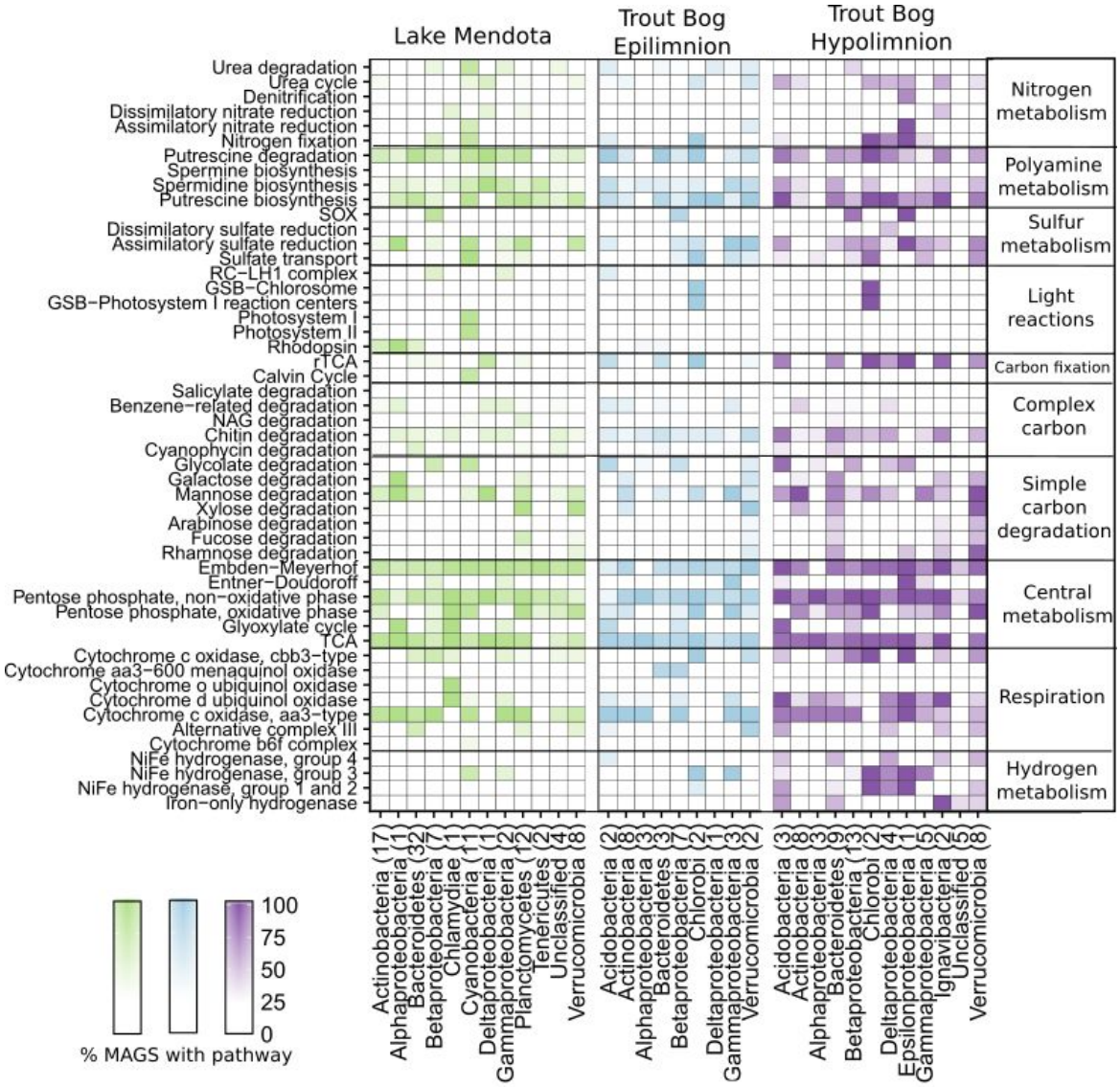

Figure 4.4. Metabolisms in Lake Mendota and Trout Bog. Metabolic pathways were predicted for all MAGs based on their gene content. At least 50% of enzymes in a pathway must have been encoded in the genome for a pathway to be considered present, as well as encoding enzymes unique to or required for a pathway. Putative pathway presence was aggregated by lake and phylum. This analysis can link potential functions identified in the metagenomes to taxonomic groups that may perform those functions. For example, MAGs with putative pathways for carbon fixation also likely fix nitrogen in both lakes. Similar, putative degradation pathways for rhamnose, fucose, and galactose were frequently encoded in the same MAGs. *Proteobacteria* was split into classes due to the high diversity of this phylum.

We observed assimilatory sulfate reduction more frequently than dissimilatory sulfate reduction, suggesting that in these populations, sulfate is more commonly used for biosynthesis, while reduced forms of sulfur are used as electron donors for energy mobilization. This is in contrast to marine systems, where sulfate reduction holds a central role as an energy source for organotrophic energy acquisition (160), although sulfate reduction could also be occurring in Lake Mendota's hypolimnion. Sulfur oxidation pathways were observed in MAGs classified as *Betaproteobacteria* from both lakes and *Epsilonproteobacteria* in the Trout Bog Hypolimnion.

Phototrophy

Primary production (the coupling of photosynthesis and carbon fixation) is a critical component of the freshwater carbon cycle. To identify differences in routes of primary production between freshwater environments, we compared marker genes for carbon fixation across sites. RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase), the marker gene for carbon fixation via the Calvin-Benson-Bassham (CBB) pathway, was most frequently observed in the epilimnion of Trout Bog (Figure 4.1). In contrast, citrate lyase, the marker gene for the reverse TCA cycle, was observed most frequently in Trout Bog's hypolimnion.

We next assessed the MAGs for photoautotrophy, expecting to find differences between our two study sites based on the observed contrasts in the functional marker

gene analysis (Figure 4.4). In Lake Mendota, the majority of MAGs encoding phototrophic pathways were classified as *Cyanobacteria*. These populations contained genes encoding enzymes in the CBB pathway. In Trout Bog, most MAGs encoding phototrophy were classified as *Chlorobium clathratiforme*, a species of *Chlorobiales* widespread in humic lakes (161). The Chlorobiales MAGs in Trout Bog contained genes encoding citrate lyase and other key enzymes in the reductive tricarboxylic acid (TCA) cycle, an alternative carbon fixation method commonly found in green sulfur bacteria such as *Chlorobi (162, 163)*. As *Chlorobium* is a strictly anaerobic lineage, the presence of citrate lyase in these populations may explain why this gene was observed more frequently in metagenomes from Trout Bog's hypolimnion. These photoautotrophs from both lakes also contained genes potentially encoding nitrogen fixation. The co-occurrence of fixation pathways in these populations are especially interesting given their relatively high abundance in their respective lakes.

The reductive TCA cycle is the only carbon fixation pathway known to be active in cultured representatives of *Chlorobiales*, but we found genes annotated as the RuBisCO large subunit (*rbcL*) were observed in some of the *Chlorobiales* MAGs. Homologs of *rbcL* have been previously identified in isolates of *Chlorobium,* and were associated with sulfur metabolism and oxidative stress (164). Inspection of the neighborhoods of genes annotated as *rbcL* in the *Chlorobiales* MAGs revealed genes putatively related to rhamnose utilization, LPS assembly, and alcohol dehydrogenation, but no other CBB

pathway enzymes. Given this information, it seems likely that this *rbcL* homolog encodes a function other than carbon fixation in the *Chlorobiales* MAGs.

The potential for photoheterotrophy via the aerobic anoxygenic phototrophic pathway was identified in several MAGs from all lake environments, especially from epilimnia, based on the presence of genes annotated as *pufABCLMX, puhA,* and *pucAB* encoding the core reaction center RC-LH1 (91). *Betaproteobacteria* and *Gammaproteobacteria*, particularly MAGs classified as *Burkholderiales*, most often contained these genes, although they were not broadly shared across the phylum (Figure 4.4). As aerobic anoxygenic phototrophy has previously been associated with freshwater *Proteobacteria* (91), these results are not surprising. Unexpectedly, an *Acidobacteria* MAG from the Trout Bog epilimnion also contained genes suggesting aerobic anoxygenic phototrophy.

Another form of photoheterotrophy previously identified in freshwater is the use of light-activated proteins such as rhodopsins (91). We observed genes encoding rhodopsins in MAGs from each lake environment, but more frequently in *Actinobacteria* and *Bacteroidetes* MAGs from Lake Mendota (Figure 4.4). Trout Bog, especially the hypolimnion, harbored fewer, less diverse MAGs encoding rhodopsins than those from Lake Mendota.

Complex Carbon Degradation

Biopolymers in freshwater can be either autochthonous (produced within the lake, ex. algal polysaccharides) or allochthonous (imported from the surrounding landscape, ex. cellulose). Organic carbon in freshwater is often classified as either autochthonous or allochthonous carbon, but this distinction has little relevance for organotrophic bacteria. For example, there is substantial overlap in the molecular composition of algal exudates, cellulose degradation intermediates, and photochemical degradation products (165, 166). One-carbon compounds such as methane are produced in the lake (therefore autochthonous), but they are also produced from decomposition of allochthonous carbon. We therefore found it more informative to categorize the carbon degradation pathways observed in our dataset by type of metabolism rather than carbon origin.

Degradation of high-complexity, recalcitrant carbon compounds requires specialized enzymes, but a wide availability of these compounds can make complex carbon degradation an advantageous trait. One way to predict the ability to degrade high-complexity carbon in microbial populations is by identifying genes annotated as glycoside hydrolases (GHs), which encode enzymes that break the glycosidic bonds found in complex carbohydrates. A previous study of *Verrucomicrobia* MAGs from our dataset found that the profiles of GHs differed between Lake Mendota and Trout Bog, potentially reflecting the differences in available carbon sources (35). Here, we expanded this analysis of glycoside hydrolases to all of the MAGs in our dataset to

identify differences in how populations from our two study sites degrade complex carbohydrates.

We calculated the coding density of GHs, defined as the percentage of coding regions in a MAG annotated as a GH to identify differences in carbon metabolism between MAGs from different lake environments (Figure 4.5). Our GH coding density metric was significantly correlated with the diversity of GHs identified (r^2= 0.39, p = $4.5x10^{-8}$), which is an indicator of the number of substrates an organism can utilize. The MAGs with the highest GH coding densities were classified as *Bacteroidales, Ignavibacteriales, Sphingobacteriales*, and *Verrucomicrobiales* from Trout Bog's hypolimnion. Two of these orders, *Sphingobacteriales* and *Verrucomicrobiales*, also contained MAGs with high GH coding densities in Lake Mendota and Trout Bog's epilimnion. There were several additional orders with high GH coding density that were unique to Lake Mendota, including *Mycoplasmatales (Tenericutes), Cytophagales (Bacteroidetes), Planctomycetales (Planctomycetes),* and *Puniceicoccales (Verrucomicrobia).* In concordance with their ability to hydrolytically degrade biopolymers to sugars, MAGs with high GH coding densities also contained putative degradation pathways for a variety of sugars (Figure 4.4).

We identified genes encoding for several GH families in MAGs from all lake environments. Starting with the most frequently observed in MAGS from all sites, these included GH109 (alpha-N-acetylgalactosaminidase), GH74 (endoglucanase), and GH23 (soluble lytic transglycosylase). However, previous research has shown that abundance

of genes annotated as GH families may be misleading (35); therefore, we prefer not to speculate on the relative importance of GH family annotations in our MAGs based on observation frequency. Lake Mendota contained unique GHs belonging to the family GH13 (alpha-glucoside). The only unique GH found in Trout Bog's epilimnion was GH62, a putative arabinofuranosidase. Trout Bog's hypolimnion contained many more unique enzymes, the most abundant of which were GH129 (alpha-N-acetylgalactosaminidase), GH89 (alpha-N-acetylglucosaminidase), GH43_12 (xylosidase/arabinosidase), GH44 (beta-mannanase/endo-beta-1,4-glucanase), GH66 (dextranase), and GH67 (alpha-glucuronidase). The increased diversity of these genes found in Trout Bog's hypolimnion suggests differences between the GH profiles, which could be correlated to differing diversity and complexity of the available organic carbon.

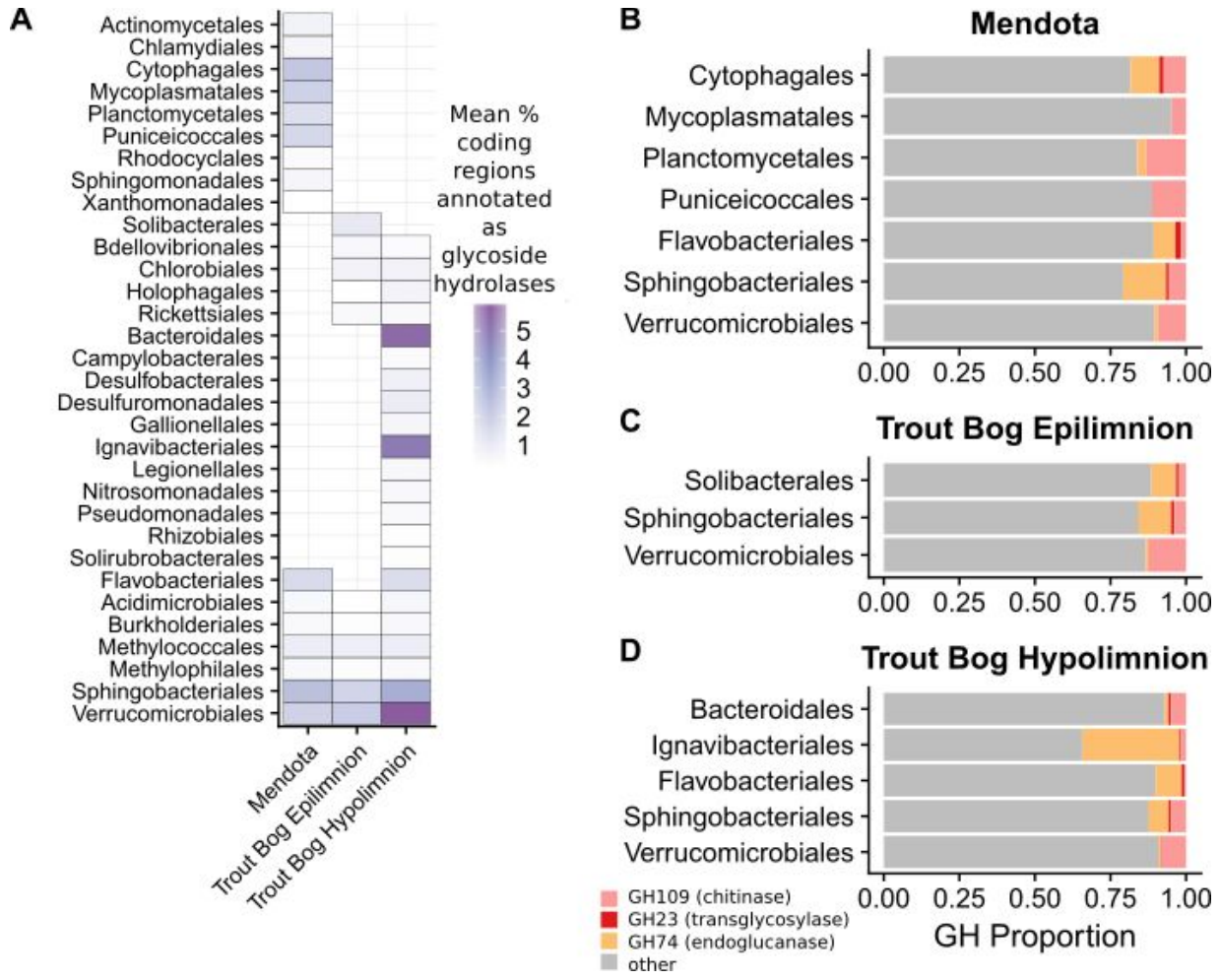

Figure 4.5. Glycoside hydrolase content in the MAGs. Annotations of GHs were used as an indication of complex carbon degradation. Genes potentially encoding GHs were identified and assigned CAZyme annotations using dbCAN. GH coding density was calculated for each MAG and averaged by order and lake (A). While a few orders contained genes encoding glycoside hydrolases in all three sites, many orders were unique to each site. The orders with the highest coding density were all found in the Trout Bog hypolimnion. Glycoside hydrolase diversity, an indicator of the range of substrates an organism can degrade, was significantly correlated with coding density ($r^2 = 0.38$, $p = 4.5 \times 10^{-8}$). Within MAGs with high glycoside hydrolase density, three families appeared most frequently - GH74, GH109, and GH23, although these abundances may be method-dependent (He et al., 2017) (B-D). *Proteobacteria* was split into classes due to the high diversity of this phylum.

Central Metabolism and Simple Carbon Degradation

Freshwater microbes are exposed to a great variety of low-complexity carbon sources such as carbohydrates, carboxylic acids, and one-carbon (C1) compounds. The central metabolic pathways shared by most living cells are often an entry point for the least complex carbon compounds. The specific routing of central metabolism may therefore reveal how low complexity carbon compounds are used. Genes encoding enzymes in the glyoxylate cycle, a truncated version of the TCA cycle that is used to produce biosynthetic intermediates and bypass decarboxylation steps, were observed in *Alphaproteobacteria* and *Chlamydiae* in Lake Mendota and *Acidobacteria* and *Betaproteobacteria* in Trout Bog. This may indicate an adaptation to reduce carbon demand in these populations.

Oxidative phosphorylation is an important part of central metabolism for aerobic bacteria, so we investigated the types of cytochrome oxidases encoded in our MAGs (Figure 4.4). Cytochrome c oxidases, both aa3- and cbb3-type, were widespread in all three lake environments and frequently co-occurred within MAGs. aa3-type cytochromes are associated with high oxygen concentrations and cbb3-type cytochromes are associated with low oxygen concentrations (167), so the presence of genes encoding both types suggests the flexibility to operate under a range of oxygen concentrations. Of the quinol-based cytochrome oxidases, genes encoding cytochrome d oxidase were most often observed in MAGs from Trout Bog's hypolimnion, while cytochrome aa3-600 was found only in MAGs classified as *Bacteroidetes* and

Betaproteobacteria from Trout Bog's epilimnion. Cytochrome o oxidase was observed only in a *Chlamydia* MAG from Lake Mendota. Alternative complex III was identified in MAGs of *Verrucomicrobia* in all sites, in *Acidobacteria* from Trout Bog (both layers), and in *Bacteroidetes* and *Planctomycetes* from Lake Mendota.

Similarly, hydrogen metabolism can influence other aspects of a microbe's nutrient usage. Iron-only hydrogenases were found primarily in MAGs from Trout Bog's hypolimnion (Table 4.3), consistent with their previously identified presence in anaerobic, often fermentative bacteria (168) and the higher observations of marker genes for iron-only hydrogenases in the hypolimnion site. Genes encoding [Ni-Fe] hydrogenases of groups 1 and 2, involved in hydrogen uptake, sensing, and nitrogen fixation, were found at significantly different frequency in all sites with the exceptions of group 2a in Lake Mendota and Trout Bog's epilimnion and group 2b in both layers of Trout Bog. Genes encoding these hydrogenases were widespread in MAGs from Trout Bog's hypolimion, found only in *Chlorobiales* MAGs in Trout Bog's epilimnion, and rarely observed in MAGs from Lake Mendota. Group 3 [Ni-Fe] hydrogenases were detected differentially at each site dependent on their subtype and were identified in MAGs belonging to *Cyanobacteria* and *Chlorobiales* in both lakes. This finding is consistent with the proposed function of Group 3d, which is to remove excess electrons produced by photosynthesis. Group 4 [Ni-Fe] hydrogenases were not observed significantly more or less in any site.

Low molecular weight carbohydrates such as glucose, fucose, rhamnose, arabinose, galactose, mannose, and xylose may be derived either from algae or from cellulose degradation (166, 169). To understand how these compounds are used by freshwater populations, we analyzed putative sugar degradation pathways in our MAGs. Genes encoding the pathway for mannose degradation, which feeds into glycolysis, appeared frequently in both lakes. Genes encoding the degradation of rhamnose and fucose, whose pathways converge to enter glycolysis and produce pyruvate, were frequently found within the same MAGs (including members of *Planctomycetes* and *Verrucomicrobia* from Lake Mendota, and members of *Bacteroidetes, Ignavibacteria,* and *Verrucomicrobia* from Trout Bog). Putative pathways for galactose degradation were often observed in these same MAGs. Xylose is a freshwater sugar which has already been identified as potential carbon source for streamlined *Actinobacteria (92)*; we confirmed this in our MAGs, and found that *Bacteroidetes, Planctomycetes,* and *Verrucomicrobia* from Lake Mendota and *Bacteroidetes* and *Verrucomicrobia* from Trout Bog were additional potential xylose degraders. Genes for the degradation of glycolate, an acid produced by algae and consumed by heterotrophic bacteria (77), were identified in *Cyanobacteria* and *Betaproteobacteria* MAGs from Lake Mendota and in *Acidobacteria, Verrucomicrobia, Alpha-, Beta-, Gamma-*, and *Epsilonproteobacteria* MAGs from Trout Bog.

Methylotrophy, the ability to grow solely on C1 compounds such as methane or methanol, appears in MAGs from both Trout Bog and Lake Mendota. Putative pathways

for methanol degradation were found in MAGs classified as *Methylophilales* (now merged with *Nitrosomonadales* (170)) and *Methylotenera*, while *Methylococcales* MAGs were potential methane degraders based on the presence of genes encoding methane monooxygenase. *Methylococcales* MAGs from Trout Bog also encoded the pathway for nitrogen fixation, consistent with reports of nitrogen fixation in cultured isolates of this taxon (171). The *Methylophilales* MAGs also likely degrade methylamines, based on the presence of genes encoding the N-methylglutamate pathway or the tetrahydrofolate pathway (172). Methylotrophy in cultured freshwater isolates from these taxa is well-documented (109, 173); however, genes encoding methanol degradation were also identified in MAGs classified as *Burkholderiales* and *Rhizobiales* from Trout Bog. Given the rapid rate at which we are discovering methylotrophy in microorganisms not thought to be capable of this process, identifying potential new methylotrophs in freshwater is intriguing, but not surprising (174).

MAGs over Time

Our metagenomes comprise a time series, so we can use MAG coverage and the number of marker gene hits as proxies for abundance over time. As an example, we analyzed abundance data for *Cyanobacteria*, known to be highly variable over time in Lake Mendota (Figure 4.6, A-E). We found that one *Cyanobacteria* MAG in each year was substantially more abundant than the rest; this single MAG only is plotted for each year. Since our analysis of the diversity of MAGs containing nitrogenases showed a strong association between nitrogen fixation and *Cyanobacteria* in Lake Mendota, we

hypothesized that the number of hits to the most abundant marker genes encoding nitrogenase subunits over time would be correlated to the abundance of the most abundant *Cyanobacteria* MAG in each year (Figure 4.6, F-J). This hypothesis was partially supported. Two of the marker genes, TIGR1282 (*nifD*) and TIGR1286 (*nifK* specific for molybdenum-iron nitrogenase), correlated with the *Cyanobacteria* MAG abundance more frequently than the third, TIGR1287 (*nifH,* common among different types of nitrogenasese). Significant correlations ($p < 0.05$) were only detected in 2008, 2011, and 2012. The strength of these correlations suggests that in three out of the five years in our Lake Mendota time series, a single *Cyanobacteria* population produced most genes encoding nitrogenase subunits. In the other two years, it is possible that other diazotrophic populations were more abundant, or that the nitrogenase subunits were derived from populations that did not assemble into MAGs. These two years were also unusual in our time series - in 2008, extreme flooding events led to large *Cyanobacteria* blooms (175) and in 2009, the invasive spiny water flea population drastically increased in Lake Mendota (176). Still, our time series analysis demonstrates the utility of our datasets in linking metabolic function to specific taxonomic groups.

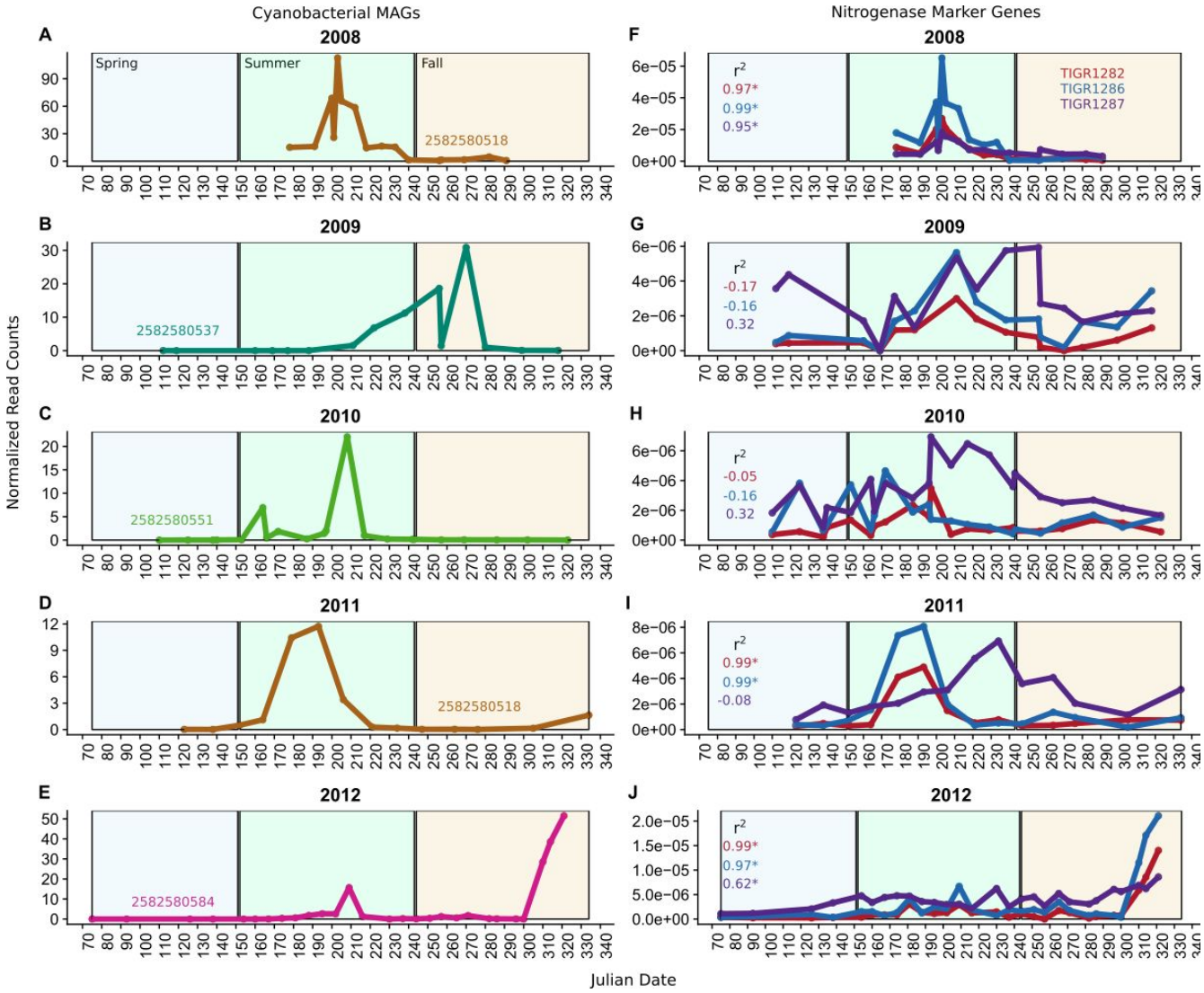

Figure 4.6. Cyanobacteria and nitrogen fixation over time. To approximate the abundance of populations over time, we mapped metagenomic reads back to the MAGs. The number of BLAST hits of marker genes in the metagenomes was used as a proxy for gene abundance. Counts were normalized by metagenome size, and in the case of the MAGs, genome length. Data from Cyanobacterial MAGs and nitrogen fixation marker genes are shown here. Colored numbers on panels A, C, E, G, and I indicate the IMG OID of the most abundant MAG in that year of data, plotted here. The marker genes used were TIGR1282, TIGR1286, and TIGR1287, encoding subunits of Mo-Fe nitrogenase; these were the most frequently observed nitrogenase markers in the Lake Mendota metagenomes. Significantly correlated trends over time were observed in the MAGs and the nitrogenase marker genes in 2008, 2011, and 2012. This suggests that nitrogen fixation is driven by these particular MAGs in those years, and is consistent with our result indicating that genes encoding nitrogen fixation were found in these MAGs. The lack of significant correlations in other years may be due to contributions from unassembled populations or more even abundances of other diazotrophic populations in that year.

Conclusions

Our analysis of functional marker genes indicated significant differences in microbial nutrient cycling between Lake Mendota's epilimnion, Trout Bog's epilimnion, and Trout Bog's hypolimnion. By combining these results with metabolic pathway prediction in MAGs, we identified taxa encoding these metabolisms and co-occurrence of pathways within MAGs. We found that phototrophy, carbon fixation, and nitrogen fixation co-occurred within the abundant phototrophs *Cyanobacteria* in Lake Mendota and *Chlorobiales* in Trout Bog. In Lake Mendota, nitrogen fixation was predominantly associated with *Cyanobacteria*, it was not associated with any particular taxon in Trout Bog. In the sulfur cycle, we observed assimilatory pathways more frequently than dissimilatory pathways in the MAGs, suggesting a bias towards using sulfur compounds for biosynthesis rather than as electron donors. We found the greatest density and diversity of genes annotated as GHs in the Trout Bog hypolimnion, potentially indicating a greater reliance on complex carbon sources in this environment. Our combination of functional marker gene analysis and MAG pathway prediction provided insight into the complex metabolisms underpinning freshwater communities and how microbial processes scale to ecosystem functions.

Chapter 5: Metatranscriptomics reveals interactions between phototrophs and heterotrophs in freshwater

Alexandra M. Linz[1], Frank O. Aylward[2], Stefan Bertilsson[3], Katherine D. McMahon[1,4]

[1]Department of Bacteriology, University of Wisconsin–Madison, [2]Department of Biological Sciences, Virginia Tech, [3]Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University, [4]Department of Civil and Environmental Engineering, University of Wisconsin–Madison

A.M. Linz designed experiments, performed experiments, analyzed data, and wrote chapter
F.O. Aylward advised data analysis
S. Bertilsson designed experiments
K.D. McMahon designed experiments and advised analysis

Introduction

Many of the core ecosystem functions of a freshwater lake, such as primary production, decomposition, and biogeochemical cycling, are driven by microbial communities. While the number of reactions performed by each cell is miniscule, the sum of all these cells is a dynamic, interconnected community whose functions can impact an entire ecosystem. Diel trends in freshwater have been documented for a number of variables that can be linked to microbial activity, including rates of primary production, nitrogen fixation, and sulfide oxidation (177–179). To better understand the impact of microbial metabolisms on freshwater ecosystems, we used metatranscriptomic sequencing to ask how microbial community functioning varies during day/night cycles in three lakes with different trophic statuses. We hypothesized that we would see diel trends in heterotrophic gene expression, driven by phototrophic gene expression, in all three lake types.

Predicting diel trends in heterotrophic microbes may seem counterintuitive. However, microbial communities are highly interconnected (72), and interactions between phototrophic and heterotrophic freshwater microbes have previously been documented. A previous metatranscriptomic study in a phosphorus-limited freshwater lake found differential gene expression in day vs. night in both phototrophs and heterotrophs, particularly in energy acquisition pathways and pyrophosphatases (180). Similar results in heterotrophic gene expression and energy acquisition pathways were also observed in a marine metatranscriptomic study (181), and correlated gene

expression between phototrophs and heterotrophs has been identified in marine systems (67). However, non-diel trends in gene expression of marine heterotrophs has also been observed (182).

Although many freshwater microbes cannot yet be cultured, co-cultures of phototrophic algae and heterotrophic bacteria can be stable. Additionally, non-cooperative interactions such as competition or predation have been observed in the laboratory (183). In the environment, changes in the phototrophic community drive variability in the heterotrophic community (184), potentially because compounds produced by the phototrophic community (such as glycolate) or decaying biomass can be consumed by heterotrophs (77). Freshwater phototrophs also release carbohydrates, which can enhance the growth of heterotrophs (169). Some ubiquitous freshwater bacteria, including *Limnohabitans,* appear to specialize in algal-derived carbon uptake (78). Abiotic photodegradation of dissolved organic carbon (DOC) could also lead to diel trends in heterotrophs (185). Given these multiple lines of evidence, we hypothesized that sunlight may be a factor driving gene expression of both light-dependent and light-independent metabolic pathways.

In this study, we use a high-resolution metatranscriptomic time series to investigate differences in community gene expression in day vs. night. We collected metatranscriptomic samples every four hours for two days and repeated this experiment in three sites representing oligotrophic, eutrophic, and humic lake types. Here, we identify a number of metabolic pathways that are differentially expressed between lakes

and between times and propose mechanisms for these observations based on the available knowledge of freshwater microbial community interactions.

Methods

Study design and in situ measurements

Three lakes in Wisconsin, USA, were chosen for this study based on their differing trophic statuses: oligotrophic (Sparkling Lake), eutrophic (Lake Mendota), and humic (Trout Bog). These lakes were chosen because they are core sites of the North Temperate Lakes - Long Term Ecological Research program, and as such, have a rich context of historical environmental data and automated sensor platforms deployed at the time of sampling. Previous studies of the resident microbial communities have also been performed in all three of these sites, providing reference genomes specific to each lake (54, 131, 186). The limnological characteristics of each lake are presented in Table 5.1.

Table 5.1. Comparison of Sparkling Lake, Lake Mendota, and Trout Bog. These three lakes were chosen for comparative metatranscriptomics because of their varying trophic statuses, extensive historical data, and previous microbial sampling. Data courtesy of NTL-LTER <www.lter.limnology.wisc.edu>. Due to thunderstorms the night of July 8th, the final 1AM timepoint in Sparkling Lake was collected on July 9th instead.

	Lake Mendota	Trout Bog	Sparkling Lake
Surface area (km2)	39.6	0.001	0.637
Maximum depth (m)	25.3	7.9	20
Trophic status	Eutrophic	Humic	Oligotrophic
Location	Madison, WI, USA	Boulder Junction, WI, USA	Boulder Junction, WI, USA
GPS coordinates	43.1113, -89.4255	46.0412, -89.6861	46.0091, -89.6695
Development on shoreline	High	Low	Moderate
pH	8.4	4.8	7.4
Total phosphorus (ug/L)	109.5	46.6	15.0
Total nitrogen (ug/L)	860	961	371
Chlorophyll (ug/L)	4.8	16.2	2.2
Secchi depth (m)	4.8	1.1	6.2
Sampling dates (2016)	July 14-16	July 8-10	July 6-9
Sunrise/sunset time on sampling dates	5:32/20:35	5:18/20:49	5:17/20:50

Each lake was sampled twelve times at four hour intervals, starting at 5:00AM and continuing until 1:00AM 44 hours later. The lakes were sampled in early July within a two week time period to minimize seasonal changes. Due to the difference in latitude, the day length at Sparkling Lake and Trout Bog was slightly longer than at Lake Mendota. Half an hour prior to each timepoint, an instrumented sonde (Hydrolab DS5X, OTT Hydromet) equipped with sensors for temperature, optical dissolved oxygen concentrations, pH, conductivity, turbidity, and chlorophyll was used to collect measurements from the top ten meters of the water column (in Trout Bog, which is only eight meters deep, the whole water column was measured). Photosynthetically active radiation (PAR) was also measured at this time using a PAR meter (Li-Cor). PAR readings were taken every half meter depth until light extinction or six meters.

Exactly at the timepoint, an integrated water sample of the epilimnion was collected. The sampling depth was chosen based on the location of the thermocline measured the day prior to beginning each lake's two day time series. The collection depth remained constant throughout sampling. All collection tools were washed with ambient epilimnion water immediately prior to each timepoint. RNA samples were the first samples collected at each timepoint. Water from the integrated epilimnion sample was pumped through 0.22 micron polyethylene filters (Supor) with a cheesecloth pre-filtration. This process occurred in the field using a Masterflex E/S portable sampler (Cole-Parmer). Each sample was filtered for the same amount of time based on the rate of filter clogging, determined for each lake prior to beginning the time series (2-4

minutes), and the volume filtered was recorded. Four replicate filters were collected. Filters containing RNA were placed in 2mL plastic cryovials (Phenix) and immediately flash frozen in liquid nitrogen. Filters were stored at -80C after collection, with the Trout Bog and Sparkling Lake samples spending four hours on dry ice during transport back to the laboratory.

After collecting RNA, additional samples were taken for lab-based measurements of environmental variables from the same epilimnion sample. Samples for total and dissolved nitrogen and phosphorus concentrations were collected in 150 mL HDPE bottles (Nalgene), with samples for dissolved nutrient analysis collected using effluent from the RNA filtration. 1L of unfiltered water for chlorophyll measurement were collected in black, opaque glass bottles and filtered on shore. Three replicate 0.3 micron nitrocellulose filters (Whatman) of 250mL were collected for chlorophyll analysis and immediately flash frozen in liquid nitrogen. 15mL of unfiltered water was collected for C14-leucine bacterial production assays in a 50 mL tube (Falcon). This sample was stored in a thermos of epilimnion water to maintain ambient temperature during transport to designated radioactive lab spaces and subsequent incubations. DNA samples for metagenomic sequencing were collected at one timepoint from each lake by filtering 250 mL of epilimnion water from the same sample used for RNA collection on to the same type of 0.22 micron filters. Cells from the water sample were preserved for single amplified genome sequencing by mixing 2mL of water with 100 uL of a

glycerol-TE buffer. Both the DNA and single cell preservation samples were flash frozen in liquid nitrogen and stored at -80C until processing.

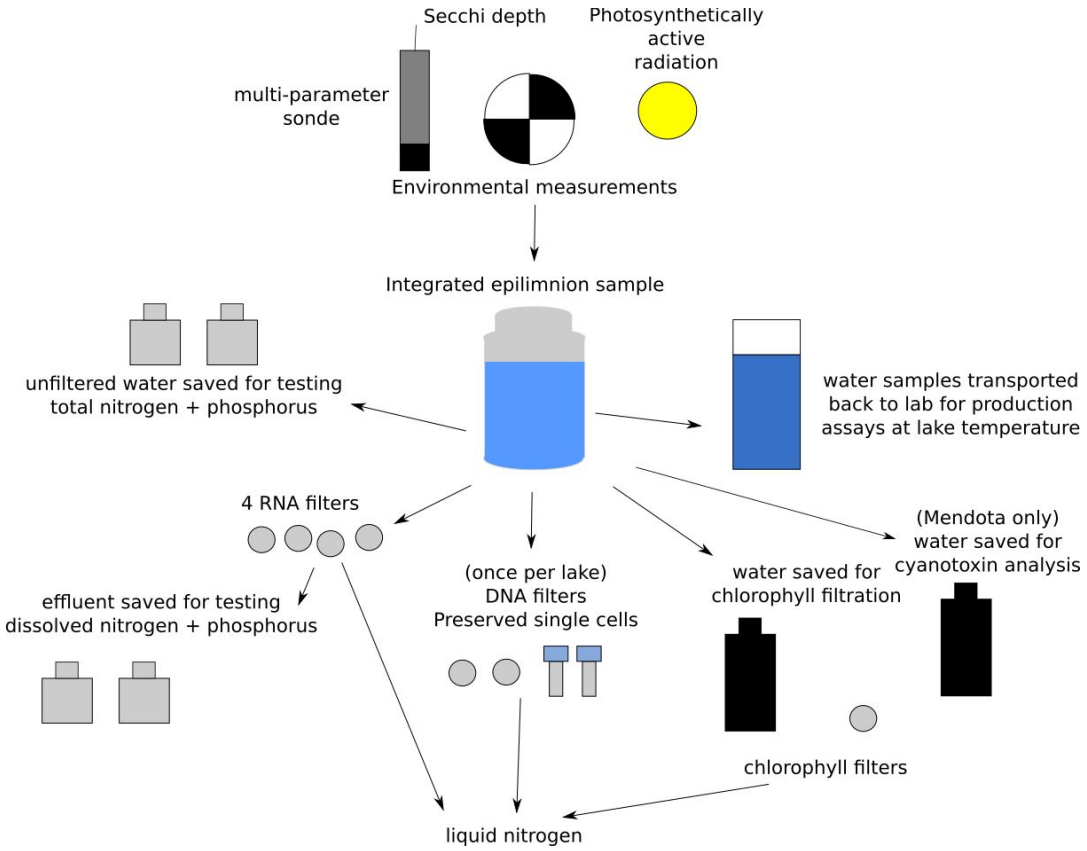

Figure 5.1. Schematic of data collected during each timepoint. After measuring qualities of the environment such as pH, light penetrations, and clarity, an integrated epilimnion sample was collected. Water from this sample was either filtered and flash frozen on the boat (for RNA), processed on shore (DNA, single cells, chlorophyll), frozen at -20C for later analysis (cyanotoxins, nitrogen and phosphorus samples) or immediately test in the lab (bacterial production assay).

RNA extraction

Within 2-3 weeks of collection, RNA was extracted from the filters in a single batch operation. A detailed protocol is available in the supplemental materials; a brief overview is presented here. Filters were exposed to a lysis solution containing EDTA and SDS and incubated at 65C. Filters were then physically destroyed using FastDNA Spin Kit reagents and bead beating protocol (MP Biomedical). TRIzol (a phenol mixture) was added to the filters before physical disruption (Thermo-Fisher). An internal standard - an *in vitro* transcription of the cloning plasmid pFN18A - was previously prepared and added to samples after beadbeating (187). The samples were centrifuged for 5 min and the supernatant was transferred to a fresh tube. From this point, the protocol resembles a typical phenol-chloroform DNA extraction. Chloroform was used to separate RNA from other molecules in the TRIzol. After cleaning, the RNA was precipitated in ethanol, pelleted, and resuspended. The RNA was further purified using an RNeasy kit (QIAGEN), which includes DNAse digestion. All samples were quantified using a Qubit fluorometer (Thermo-Fisher) and stored at -80c until sequencing. A subset of the samples were further tested on a BioAnalyzer to confirm that the RNA was of sufficient quality for sequencing (Agilent Genomics).

Additional lab-based measurements

Although the sonde used to measure characteristics of the water column included a chlorophyll sensor, sensor-based measurements of chlorophyll are highly variable. Therefore, we also took lab-based measurements of chlorophyll concentrations.

Chlorophyll was extracted from the triplicate filters using methanol following NTL-LTER protocols. Samples were acidified to measure phycocyanin in addition to chlorophyll. Extracted chlorophyll was diluted as needed to remain within range of the spectrometer; samples from Sparkling Lake required no dilution, samples from Lake Mendota required a 1:4 dilution, and samples from Trout Bog required either a 1:2 or a 1:4 dilution.

Bacterial production assays were conducted using C14-leucine at each timepoint. 1.5 mL of water were added to six microcentrifuge tubes. Two of the six samples were immediately killed using trichloroacetic acid (TCA) as negative controls. All samples received C14-leucine and were incubated for one hour, after which samples were killed with TCA and stored at -20C. Approximately one month after sample collection, production assay samples were thawed, pelleted, and resuspended in ethanol. Radioactivity was measured using a liquid scintillation counter.

DNA filters underwent a phenol/chloroform extraction using the same lysis method as the RNA extraction protocol. An additional four DNA samples collected from Sparkling Lake in 2009 were extracted and sent for sequencing to serve as additional references for this lake.

Sequencing

All samples were sequenced by the Department of Energy Joint Genome Institute (JGI). Once received, rRNA was depleted from the RNA samples. RNA samples were sequenced using Illumina HiSeq 2500-1TB. Metatranscriptomic reads were quality

filtered by JGI. Metatranscriptomic reads were assembled by JGI using MetaHit (188). DNA samples for metagenomics were also sequenced on an Illumina HiSeq platform. Metagenomic reads were assembled by JGI and assembled using MetaHit. DNA samples for 16S rRNA ribosomal gene amplicon sequencing were sequenced on an Illumina MiSeq platform. The resulting reads were filtered using BBDuk and reads mapping to human, mouse, cat, and dog genomes with BBMap were removed (189).

Cells for single amplified genomes were sorted, identified using 16S amplicon sequencing, and sequenced using JGI's standard single amplified genome (SAG) protocols. Cells for SAG sequencing were chosen with a preference towards cells from Sparkling Lake, the least well-represented lake in the dataset. An Illumina shotgun library was constructed from each single cell and sequenced on the Illumina NextSeq platform. Sequencing reads were filtered using BBTools (189) and assembled into SAGs using SPAdes (190). Unscreened SAGs were used as references to retain any unusual DNA sequences in the genome.

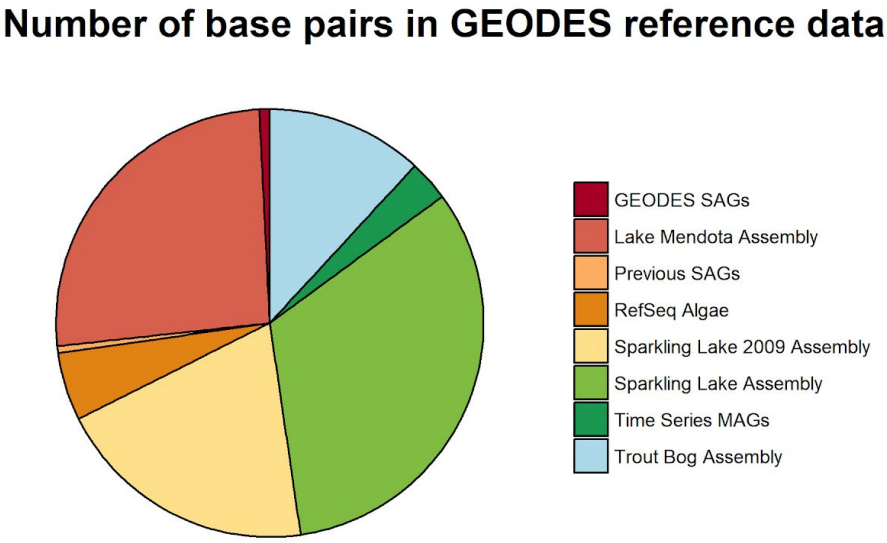

Figure 5.2. Composition of the sequences used as references for annotating and classifying metatranscriptomic reads. To allow comparisons between study sites, a single database was used for all metatranscriptomes. Sequences included in this database were a mix of newly generated metagenome assemblies and SAGs, genomes generated from previous projects at these study sites, and published reference genomes for algae. All of these sequences were clustered at 98% sequence similarity, creating a non-redundant database of coding regions. The longest gene was chosen as the representative of each cluster. After clustering, the database was reduced in size by 40%, indicating high levels of redundancy between the various input types.

Bioinformatics pipeline

Ribosomal rRNA reads, which still comprised approximately 50% of metatranscriptomic reads despite rRNA depletion, were removed using SortMeRNA (191). Assembled metagenomic contigs from this study, SAGs from this study, SAGs and MAGs from previous McMahon Lab time series sequencing on these lakes (54, 131, 186), and 5 algal genomes from NCBI RefSeq (192), representing each algal genus, were used

to build a nonredundant, highly specific database for mapping metatranscriptomic reads. This approach provides better functional prediction than annotating each individual read. After formatting each type of genome or contig's fastq and gff files, coding regions were extracted and clustered at 97% ID using CD-HIT (193). Metatranscriptomic reads were mapped to this database with a 90% ID cutoff using BBMap (189). Mapped reads were tabulated using FeatureCounts (194). Metagenome assemblies were binned by lake using Metabat and checked for completeness and contamination using CheckM (135, 146). Bins and unbinned contigs from the metagenome assemblies were classified by taking the consensus taxonomy of the best hit in the IMG database for each coding region on a contig/bin (Stevens, unpublished).

Statistics

We added an internal standard to our RNA samples immediately after cell lysis during the extraction process. This allowed both normalization using the standard instead using to library size, as is often performed, and assessment of extraction success. Samples with either too few counts of the internal standard (less than 50) or orders of magnitude higher expression in all genes after normalization when compared to replicates were discarded. After these quality control measurements, 32 samples remained from Sparkling Lake, 30 from Lake Mendota, and 21 from Trout Bog. The resulting read counts are in units of transcripts/L, and are semi-quantitative (keeping in mind the limitations and biases of metatranscriptomic sequencing).

The statistical software "R" was used for further analysis (R Core Team, 2018). To reduce noise in the dataset, the top 20,000 expressed genes in each lake were retained for further differential expression analysis. From this subset, marker genes for metabolic processes were selected and aggregated by pathway - for example, any gene annotation containing "nitrogenase," "nifH|nifD|nifK," or "nitrogen fixation" was considered representative of the pathway nitrogen fixation. The summed expression of each pathway/process was input into DESeq to test differential expression (195). Despite normalizing by the internal standard, samples were still normalized by size factors to control for compositional bias, as recommended by the authors of DESeq. To account for the large number of pathways tested, an adjusted p-value of less than 0.05 was used as an indicator of significantly different expression. This analysis was run both with lake as a condition and with day or night as a condition within each lake. Day timepoints were considered to be 9AM, 1PM, and 5PM, while night timepoints were considered to be 9PM, 1AM, and 5AM. Results were plotted using the R packages ggplot2 (Wickham, 2009) and cowplot (Wilke, 2017).

Results

What genes are expressed?

As an initial comparison between our study sites, we first asked which genes were most expressed in each lake (Table 5.2). Photosynthesis related genes, particularly those encoding the photosystem II P680 D1 protein, were highly expressed in all three lakes. Genes encoding RuBisCO, the key enzyme in carbon fixation via the

Calvin-Benson-Bassham (CBB) pathway, were among the top expressed genes in Lake Mendota and Trout Bog, but not Sparkling Lake. These genes were most frequently derived from *Cyanobacteria* or *Chitinophagia*, a member of *Bacteroidetes*. *Cyanobacteria* is a ubiquitous primary producer in freshwater, but *Chitinophagia* is not known to perform photosynthesis. As *Bacteroidetes* is proposed to form a superphylum with *Chlorobi*, it is possible that these genes are instead derived from green sulfur bacteria (196). Interestingly, a hypothetical gene from the bacterial predator *Bdellovibrio* and a gene encoding PQQ-dependent dehydrogenase were highly expressed in Lake Mendota.

Because of the dominance of phototrophic taxa and genes in the most expressed genes in all sites, we also investigated which genes were highly expressed from heterotrophic organisms (Table 5.3). Housekeeping genes such as RNA polymerase, chaperonin, and translation elongation factors were commonly expressed in all lakes. Many of the most highly expressed heterotrophic genes in Lake Mendota were classified as belonging to acI, including a sugar transporter. In Trout Bog, groups such as Verrucomicrobia and Armatimonadetes contributed some of the top expressed genes, while in Sparkling Lake, genes expressed by Deltaproteobacteria were frequently observed. Genes encoding flagellin were among the most highly expressed genes in Sparkling Lake.

Table 5.2 Top 10 most expressed genes in each study site. The top 10 most expressed genes from each lake, without any filters are presented here. Annotations and classifications are derived from the sequence to which each read mapped.

Lake Mendota	Trout Bog	Sparkling Lake
Photosystem II P680 reaction center D1 protein *Cyanobacteria*	Photosystem II P680 reaction center D1 protein *Bacteria*	Hypothetical *Unclassified*
Photosystem II P680 reaction center D1 protein *Bacteria*	RubisCo large chain *Bacteria*	Hypothetical *Unclassified*
Photosystem II P680 reaction center D1 protein *Cyanobacteria*	Photosystem II CP43 chlorophyll apoprotein *Bacteria*	Photosystem II P680 reaction center D1 protein *Cyanobacteria*
Hypothetical *Unclassified*	Putative beta-barrel porin-2 *Unclassified*	Photosystem II P680 reaction center D1 protein *Unclassified*
Hypothetical *Unclassified*	Photosystem II P680 reaction center D1 protein *Cyanobacteria*	Photosystem II P680 reaction center D1 protein *Chitinophagaceae*
RubisCo large chain *Cyanobacteria*	Photosystem I P700 chlorophyll a apoprotein A2 *Unclassified*	Hypothetical *Unclassified*
PQQ-dependent dehydrogenase *LD28*	Photosystem II CP47 chlorophyll apoprotein *Bacteria*	Hypothetical *Unclassified*
Hypothetical *Unclassified*	Hypothetical *Unclassified*	Photosystem II P680 reaction center D1 protein *Unclassified*
Hypothetical *Bdellovibrio*	Photosystem I P700 chlorophyll a apoprotein A1 *Unclassified*	Hypothetical *Unclassified*
RubisCo large chain *Cyanobacteria*	Photosystem II P680 reaction center D2 protein *Bacteria*	Hypothetical *Bacteria*

Table 5.2 Top 10 most expressed annotated genes from heterotrophs in each study site. The top 10 most expressed genes from each lake, filtered to exclude photosynthetic genes, phototrophic organisms, hypothetical genes, and unclassified genes, are presented here. Annotations and classifications are derived from the sequence to which each read mapped.

Lake Mendota	Trout Bog	Sparkling Lake
PQQ-dependent dehydrogenase *LD28*	Ig-like protein group1 *acI-B*	Chaperonin GroEL *Deltaproteobacteria*
Translation elongation factor TU *Bacteroidetes*	S-layer homology domain *Fimbriimonas*	S-layer homology domain *Fimbriimonas ginsengisoli*
YTV protein *Actinobacteria*	Ribonucleoside-diphosphate reductase alpha chain *Pedosphaera parvula*	F-type H+-transporting ATPase subunit beta *Chitinophagia*
Translation elongation factor EF-G *acI-A6*	Type II secretion, pseudopilin PulG *Verrucomicrobia*	Heat shock protein 5 *Deltaproteobacteria*
DNA-directed RNA polymerase *acI-A6*	Porin *Rhodospirillales*	Flagellin *Comamonadaceae*
Probable sodium:solute transporter *LD12*	DNA-directed RNA polymerase subunit beta *Pedosphaera*	Chaperone DnaK *Deltaproteobacteria*
DNA-directed RNA polymerase *acI-A6*	Chaperonin GroL *Bdellovibrionales*	Major capsid protein Gp23 *Candidatus Saccharibacteria (TM7)*
ABC-type sugar transport system *acI-B1*	Elongation factor Tu *Verrucomicrobia*	F-type H+-transporting ATPase subunit beta *Chitinophagia*
Translation elongation factor TU *acI-A6*	Outer membrane receptor, mostly Fe transport *Proteobacteria*	Outer membrane receptor, mostly Fe transport *Caulobacteraceae*
Por secretion system C-terminal sorting domain *Sphingobacteriia*	Prepilin-type N-terminal cleavage protein *Armatimonadetes*	Flagellin *Deltaproteobacteria*

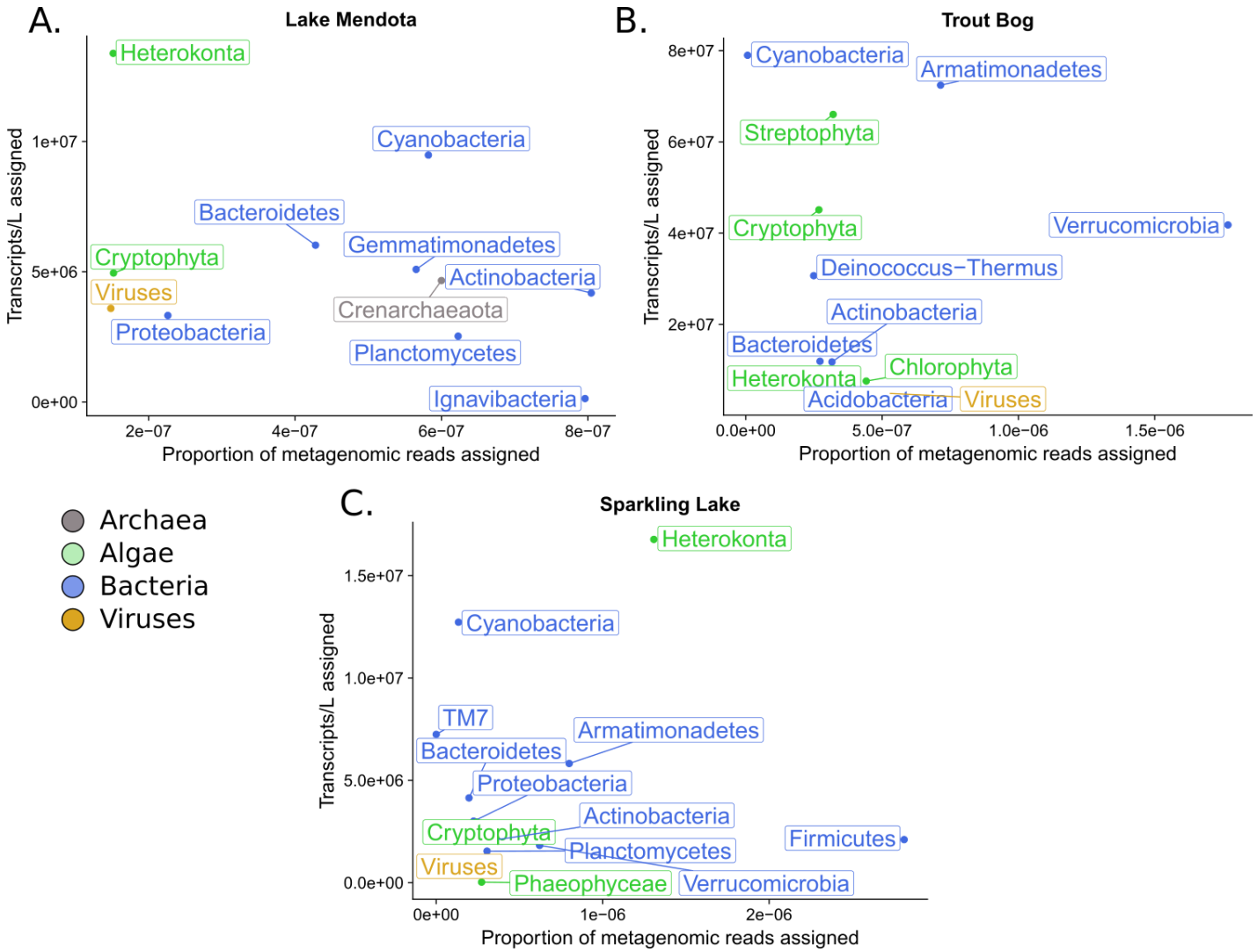

Figure 5.3 Most highly expressed or abundant phyla by lake. The expression of genes were aggregated by phylum and compared to the coverage of those phyla in metagenomes. No positive relationship was observed between expression and abundance. The identity of the most expressed and most abundant phyla varied by lake. One phylum, Chloroflexi, was removed from the plot of Lake Mendota due to orders of magnitude higher expression and abundance. This phylum is likely an outlier.

What phyla are expressed?

We next aggregated expressed genes by phylum to compare the most expressed taxa to the most abundant taxa based on metagenomic data (Figure 5.3). The same reference database was used for mapping metatranscriptomic and metagenomic data, making such comparisons possible. No positive trend between expression and abundance was observed. Eukaryotic algae were among the most expressed and most abundant phyla in all three lakes, with more types of algae observed in Trout Bog. *Cyanobacteria* were highly expressed in all three lakes, while viruses were present and expressing at low levels in all sites. The only abundant and expressing Archaeal phylum observed, *Crenarchaeota,* was detected in Lake Mendota. One phylum, *Chloroflexi*, had orders of magnitude higher expression and abundance than other phyla in Lake Mendota. This phylum is likely an outlier - genes with this classification were almost exclusively derived from a single, low quality MAG.

Trends in environmental variables

We collected data on many other environmental variables to compare these trends to those observed in gene expression, expecting that several of these trends would be diel. Parameters that reflect the boundaries between layers within the water column, such as dissolved oxygen, temperature, pH, and conductivity, were strongly diel in Lake Mendota, but less so in Sparkling Lake and Trout Bog (Figure 5.4). Concentrations of chlorophyll, often used as an indicator of primary production, were diel in Trout Bog, but not in the other two sites. Bacterial production, measured via

C14-leucine incorporation, showed dynamics over the two day time series all three lakes, although the trends were not diel (Figure 5.5). No trends were observed in nitrogen or phosphorus concentrations. It is unclear based on our metatranscriptomic data which taxa or genes may be driving trends in our measured environmental variables.

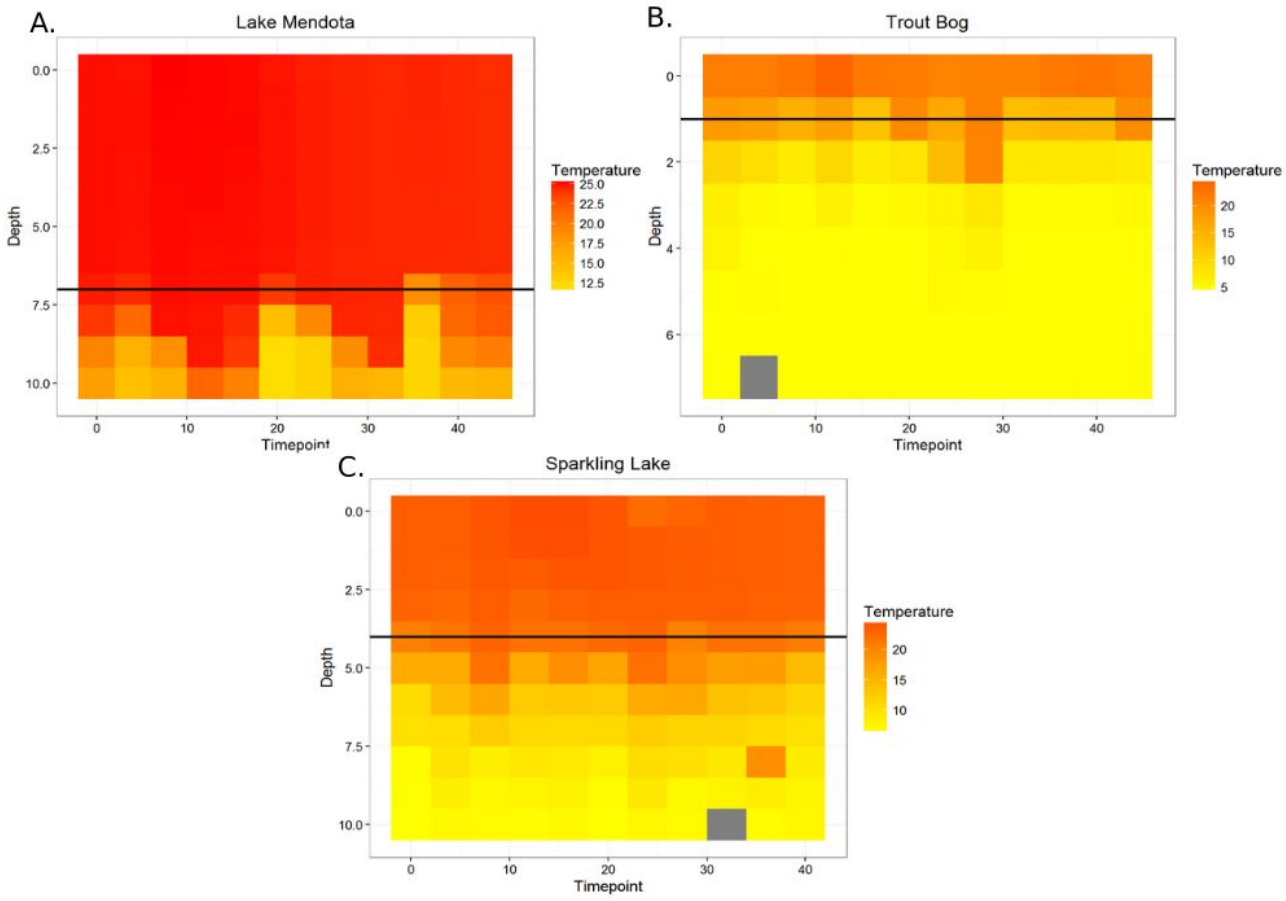

Figure 5.4. Water column temperature over the two day time series. Temperature and dissolved oxygen mark the boundaries between layers in lakes, and have previously been shown to fluctuate on diel scales. We measured these variables using a multi-parameter sonde at each timepoint. Temperature and dissolved oxygen co-varied across all lakes; we present only the temperature data here. The x-axis indicates time since the first sample was collected, and the y-axis indicates the depth at which the measurement was taken. Temperature is in Celsius, and the black line indicates the lowest depth of the integrated water column taken for RNA filtration. Temperature shows strong diel trends in Lake Mendota, but not in the other lakes.

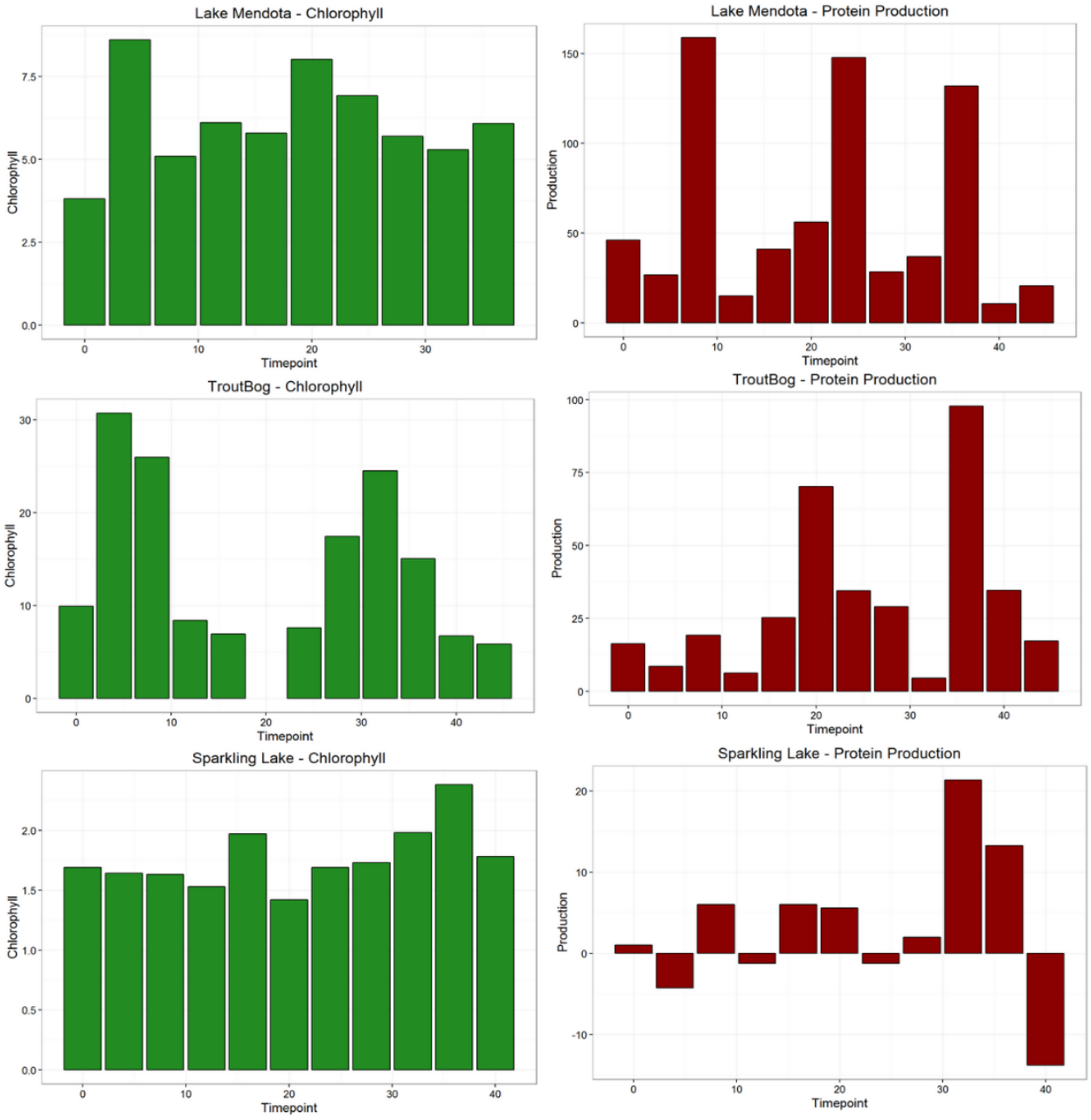

Figure 5.5. Chlorophyll concentrations and bacterial production. Filtered chlorophyll was collected at each timepoint, and bacterial production assays were used to measure protein production. Only chlorophyll concentrations in Trout Bog showed a diel trend. Both chlorophyll concentration and protein production were overall lower in Sparkling Lake than in the other two lakes.

Differential expression between lakes

We next asked how the expression of genes associated with specific functions varied by lake (Figure 5.6). The most highly expressed functional groups were genes encoding photosystem I, photosystem II, and ribulose-bisphosphate carboxylase (RuBisCO). Trout Bog had significantly higher expression of all three of these groups than Sparkling Lake and Lake Mendota. As expected, many of these reads were classified as *Cyanobacteria* in all lakes, but *Chitinophagia* was also a large contributor of genes related to primary production, especially in Sparkling Lake. In general, genes related to complex carbon degradation (hexosaminidases and glycosyl hydrolases) and one-carbon (C1) compound degradation (methane and methanol degradation) were most expressed in Trout Bog. Groups expressing complex carbon degradation genes in Trout Bog included *Verrucomicrobia, Armatimonadetes,* and *Bacteroidetes,* while C1 genes were derived from *Alphaproteobacteria, Gammaproteobacteria*, *Betaproteobacteria, Armatimonadetes, and Bacteroidetes*. Carboxylate transport was also most expressed in Trout Bog. Chitinase and chitobiose were more expressed in Lake Mendota (*Verrucomicrobia* and *Cyanobacteria)* and Sparkling Lake (*Actinobacteria* and *Burkholderiales).*

Genes related to rhodopsin biosynthesis (*Actinobacteria* and *Bacteroidetes*) and general sugar transporters (*Actinobacteria* and *Burkholderiales*) were most expressed in Lake Mendota. Genes encoding nitrate reductase (*Eukaryota*), sulfur oxidation (*Betaproteobacteria*) and cellobiose transport (*Actinobacteria* and *Proteobacteria*)

were only expressed in Sparkling Lake. These significant differences in gene expression and taxonomy suggest different microbial metabolisms are used in each lake.

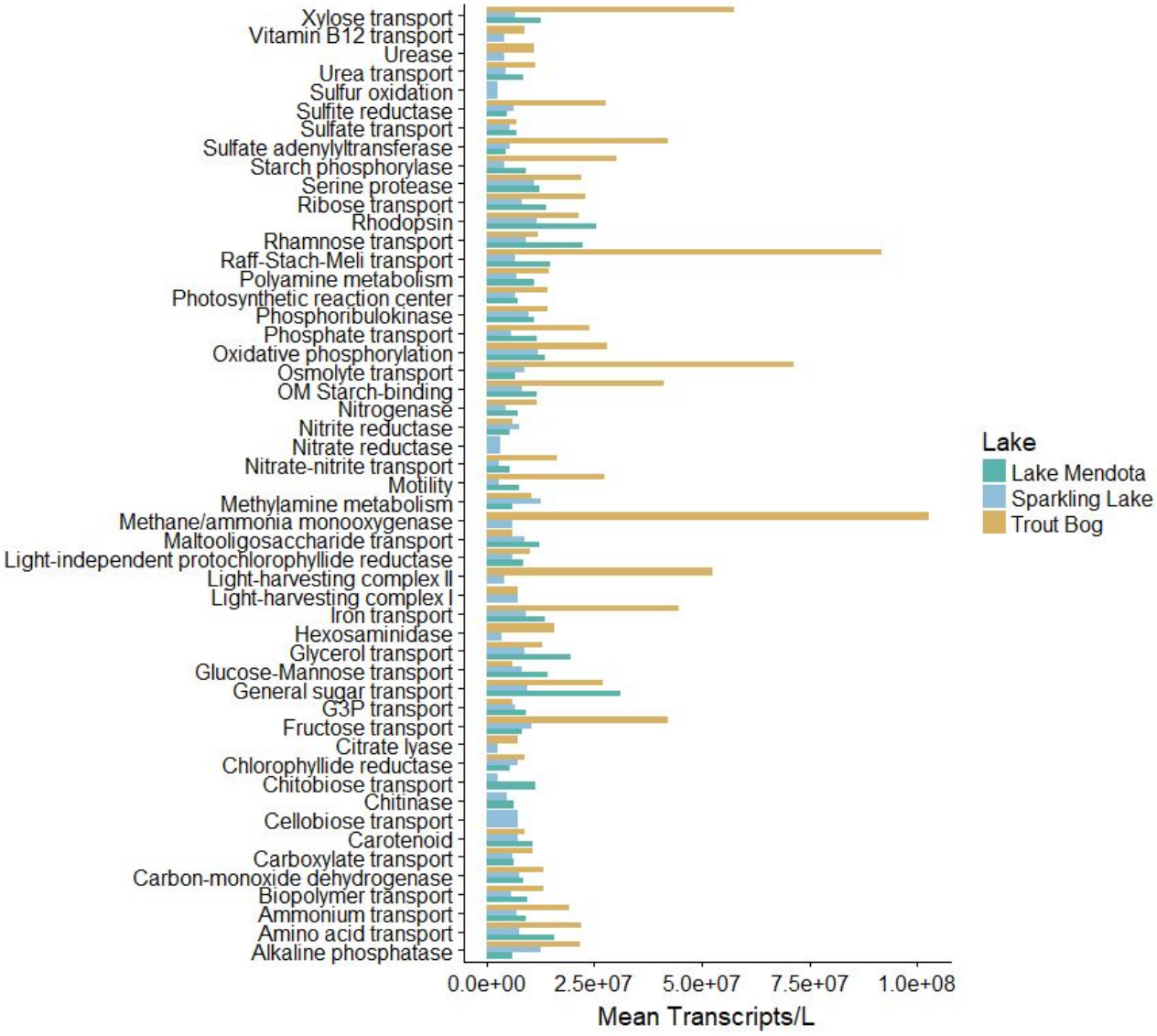

Figure 5.6. Differential expression by lake. To investigate how expression of functional groups varied by lake, we aggregated genes by associated metabolic function and tested differential expression by lake. The average expression of each group in samples from each site is shown on the x-axis. Three groups (photosystem I, photosystem II, and RuBisCO) are not shown due to orders of magnitude higher expression. In all three of these groups, expression in Trout Bog was significantly higher than in Sparkling Lake and Lake Mendota. Many other groups also differed significantly by lake.

Is there evidence of diel trends?

We designed this experiment to investigate relationships between phototrophs and heterotrophs revealed through diel trends in heterotrophic populations. Therefore, we used Fourier transformations to identify genes with significant cyclic expression over the time series in each lake. An abundance threshold was imposed on genes before testing. At a significance threshold of $p < 0.05$, we expected a false discovery rate of 5% and corrected for this. No genes in Trout Bog or Sparkling Lake were found to have significant cyclic trends after correction for false discovery; without this correction, 2.4% of genes in Sparkling Lake and 3.7% of genes in Trout Bog were determined to be cyclic at $p < 0.05$. In Lake Mendota, 7.6% of genes tested were significantly cyclic at $p < 0.05$, and 73 genes (approximately 0.01%) passed the threshold for significance after correction for false discovery. These genes were primarily from Cyanobacteria or *Bdellovibrio,* and most annotations encoded photosystem machinery.

Instead, we aggregated timepoints by day (9AM, 1PM, and 5PM) or night (9PM, 1AM, and 5AM) and tested differential expression. To reduce the number of tests, this analysis was performed on the top 20,000 most abundant genes. This analysis revealed many genes with significantly different expression in day timepoints vs. night timepoints. As expected, genes related to photosystems and oxidative phosphorylation were more highly expressed in the day than at night in all lakes. Expression in day vs. night of these genes did not partition by taxonomic classification. Sugar transporters, both of general transporters and specific substrates, were more expressed at night than

during the day in all lakes. Differential expression was not associated with taxonomic group in any lake.

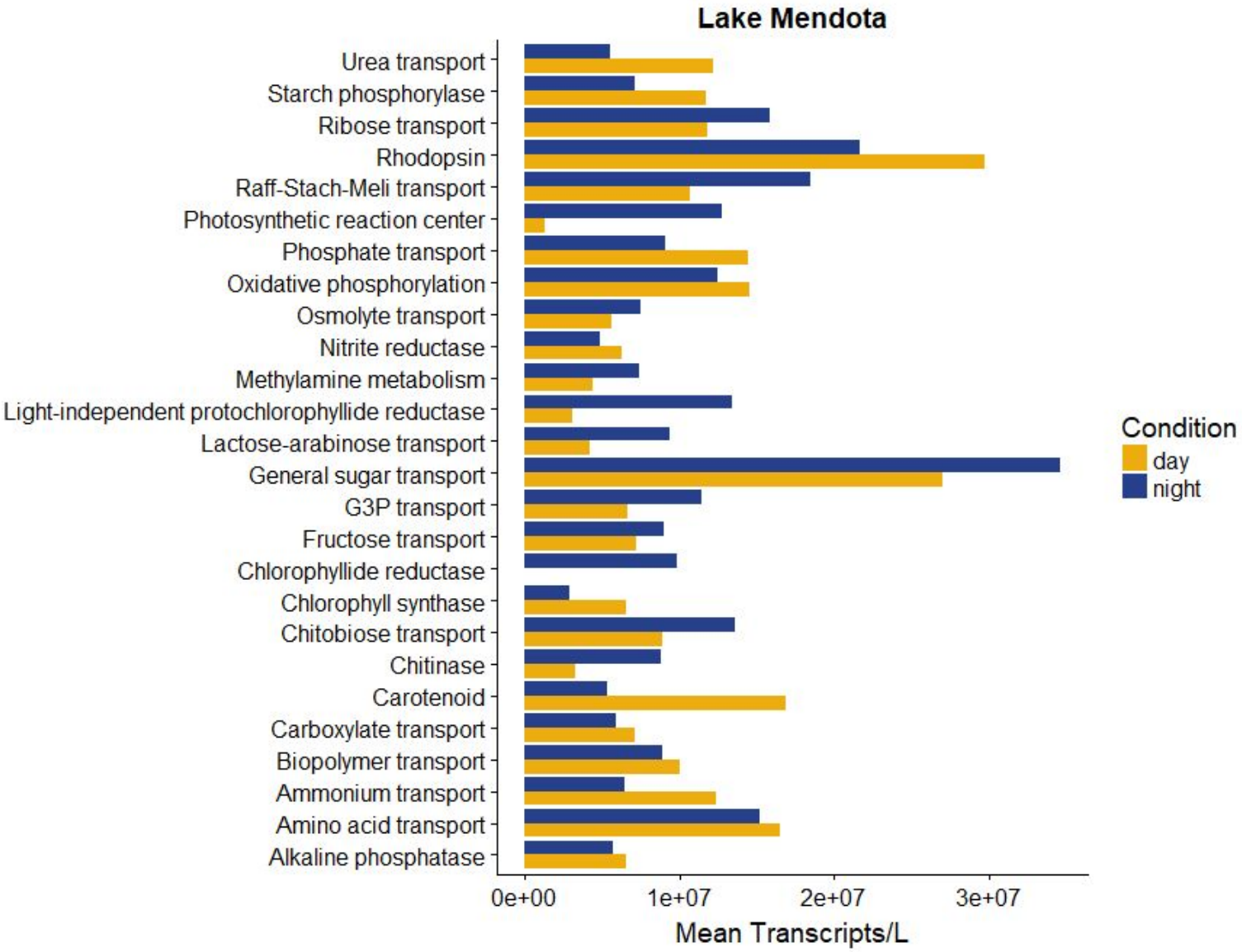

Figure 5.7. Differential expression in day vs. night in Lake Mendota. Many genes were differential expressed depending on light conditions in Lake Mendota. For example, sugar transport was more highly expressed at night, while nitrogen cycling genes were more highly expressed during the day.

Lake Mendota was the only lake were alkaline phosphatase was differentially expressed (Figure 5.7). Its expression was higher during daytime, contrary to previous research. Nitrogen-related functions were more highly expressed during day, as was rhodopsin biosynthesis. Notably, chlorophyllide reductases and photosynthetic reaction centers were more highly expressed at night. Rhodopsin-related genes showed higher expression in the day time, as did genes encoding carboxylate transporters.

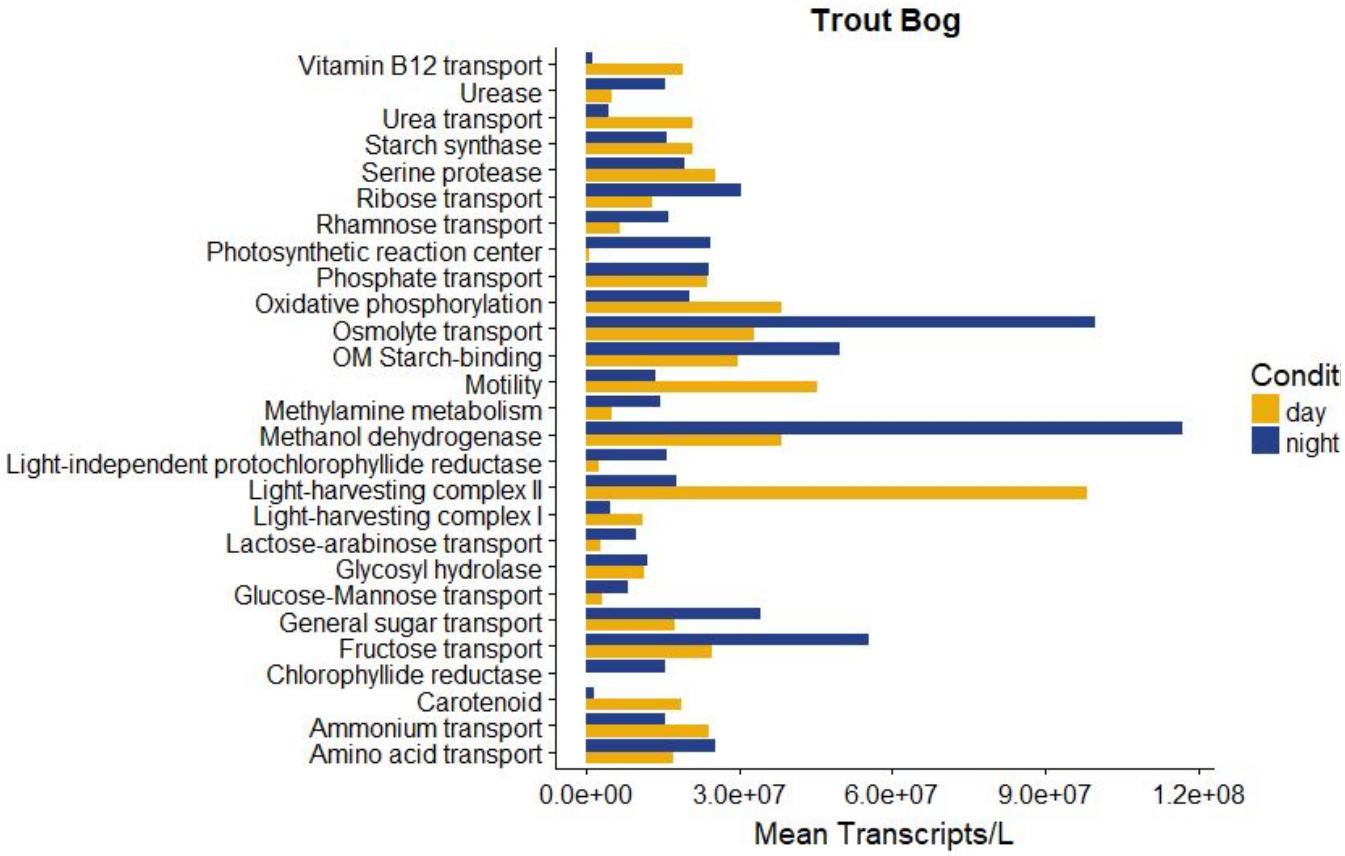

Figure 5.8. Differential expression in day vs. night in Trout Bog. In this lake, C1 enzymes were expressed more at night than during the day. Motility-related genes showed the opposite trend. Light-harvesting complex were more expressed in day, while chlorophyllide reductases and photosynthetic reaction centers were more expressed at night.

Trout Bog had several unique metabolisms that were differentially expressed in day vs. night compared to the other two lakes (Figure 5.8). Genes related to motility were more expressed during the day, while genes encoding C1-related enzymes were more expressed at night. Similar to Lake Mendota, chlorophyllide reductases and photosynthetic reaction centers were more expressed at night, although light-harvesting complex were differentially expressed and higher in the day time in Trout Bog. Rhodopsins and carboxylate transporters were not differentially expressed.

Sparkling Lake was the only lake where nitrogenase was differentially expressed; its expression was higher during the day, as was expression of nitrite reductase (Figure 5.9). As in the other two lakes, chlorophyllide reductases and photosynthetic reaction centers were more highly expressed at night, while light-harvesting complexes were more highly expressed during day. Genes encoding chitinase were more expressed in the day, while genes encoding chitobiose transporters were more expressed at night. Carboxylate transporters were more expressed in daylight. Interestingly, rhodopsin-related genes were more expressed at night than during the day.

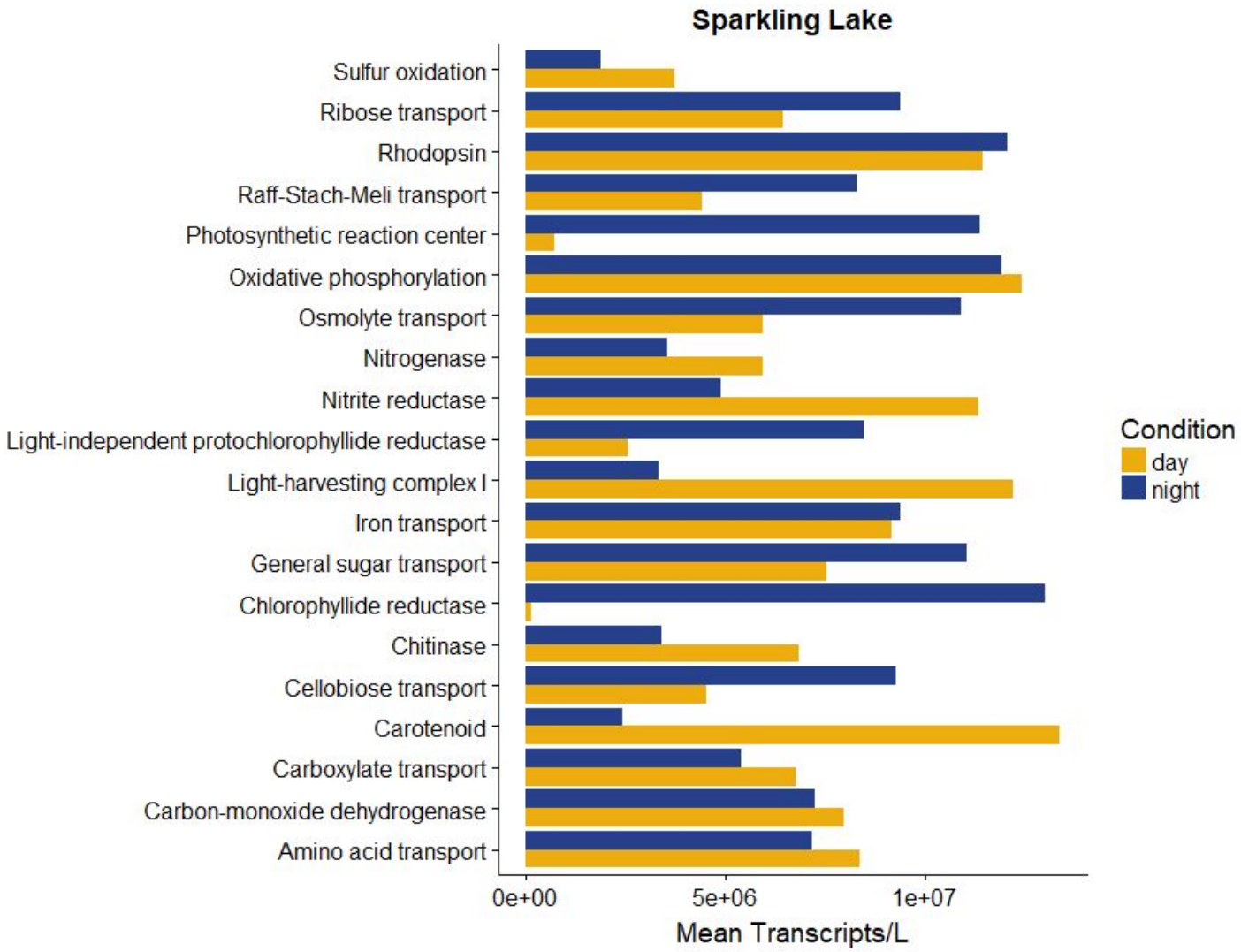

Figure 5.9. Differential expression in day vs. night in Sparkling Lake. Results of photosynthesis-related genes were largely consistent with the trends in the other two lakes, while rhodopsin-related genes were unexpectedly more expressed at night. Nitrogen cycling functions (nitrogenase and nitrite reductase) were more highly expressed during the day. Chitinase was more expressed in daytime, while cellobiose transport was more expressed at night.

Discussion

We hypothesized that interactions between phototrophs and heterotrophs in freshwater microbial communities would result in diel trends in heterotrophic gene expression. To make our findings more generalizable, we repeated this experiment in three different types of lakes. We used lake-specific reference genomes and contigs, avoiding the issues of read-based annotation, to understand how community functioning varies by lake and by time of day.

Our main research question was how gene expression varies in day vs. night in each lake. As expected based on previous metatranscriptomic studies, genes related to photosystems and oxidative phosphorylation were more expressed during daylight in all lakes (180, 181). Transports of di-and tri-carboxylates, known to be photodegradation products of dissolved organic carbon, were more highly expressed in daylight (197). These findings are consistent with our understanding of microbial metabolisms in freshwater.

Based on the documented exchange of metabolites such as carbohydrates and glycolate between phototrophs and heterotrophs, we hypothesized that expression of genes related to uptake and degradation of sugars would peak at the same time as genes related to primary production. However, the exact opposite was true. Uniformly across lakes and multiple transporter types, genes encoding sugar transport were significantly more expressed at night than during the day, opposite to the expression patterns observed in genes related to primary production. Still, the diurnal changes in sugar

uptake suggest that carbon is being passed from the phototrophic community to the heterotrophic community.

Sugar degradation at night is remarkably similar to how these metabolisms are partitioned in plants and algae: photosynthesis fuels carbon fixation during the day and the plant stores that carbon as starch, which is broken into carbohydrates and used to fuel respiration and biomass production at night (198, 199). We searched for genes encoding starch synthase and starch phosphorylase in our metatranscriptomic dataset, and we found that these genes had low expression and were primarily derived from Bacteroidetes and Cyanobacteria. This does not mean that starch accumulation and degradation is not occurring in our lakes. Rather, little to no expression of these genes may be due to the previously observed disconnect between expression levels and protein abundance (65) or to the lack of representative algal genomes from our lakes.

Algae have long been known to excrete molecules such as polysaccharides, amino acids, and even cytotoxins to their environment (200). Up to 25% of fixed carbon can be excreted by algae (201). This carbon can then be assimilated by heterotrophs (202). However, there has been debate about whether actively growing algae exude carbon or whether these compounds are derived decaying algae (203). Algal release of cytotoxins would presumably improve fitness by reducing competition, making it unlikely that these compounds are decay products (204). The marine phototroph *Prochlorococcus* likely exudates carbon to maintain redox, as it generates more reducing power than it can use via photosynthesis (205). However, a frequently observed adaptation to excess

reducing power is to downregulate photosynthesis electron flux, which is not observed in *Prochlorococcus*. The heterotrophic community is highly dependent upon these *Prochlorococcus* exudates and likely performs a community function in return, much like the co-dependency between chloroplasts and mitochondria in plant cells. In coral reefs, algal exudates can dramatically shift bacterial community composition, potentially providing algae with a competitive advantage over coral by selecting for coral pathogens in the heterotrophic community (206).

It is therefore not outlandish to hypothesize that freshwater algae may be releasing carbohydrates to change the heterotrophic community for their own benefit. Perhaps heterotrophs produce compounds that are beneficial for phototrophs, such as peroxidases or catalases, vitamins, antimicrobial peptides and antibiotics, or inorganic nutrients, all of which were expressed in our metatranscriptomic dataset. However, further experimentation is needed to confirm these hypotheses.

The research presented here asks only a handful of questions of this large metatranscriptomic dataset. Focusing on a specific lake, subset of taxa, or metabolism type also has the potential to inform our understanding of freshwater microbial functions. This dataset also informs us about the daily variability of metatranscriptomes in freshwater and the depth of sequencing required to observe trends, which is crucial knowledge when planning experiments to assess seasonal or regional trends in gene expression. All data and code from this research is publicly available at <https://github.com/alexlinz/geodes>, and we hope that this dataset will be used as a

resource by other scientists for their own questions about freshwater microbial communities.

Chapter 6: Differential expression in closely related freshwater taxa

Alexandra M. Linz[1], Frank O. Aylward[2], Stefan Bertilsson[3], Katherine D. McMahon[1,4]

[1]Department of Bacteriology, University of Wisconsin–Madison, [2]Department of Biological Sciences, Virginia Tech, [3]Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University, [4]Department of Civil and Environmental Engineering, University of Wisconsin–Madison

A.M. Linz designed experiments, performed experiments, analyzed data, and wrote chapter
F.O. Aylward advised data analysis
S. Bertilsson designed experiments
K.D. McMahon designed experiments and advised analysis

Introduction

A major question in microbial ecology is how microbial species arise and diverge. Because of their high rate of mutations and fast generation times, speciation in bacteria can be observed on the scale of months to years. Long-term evolution studies have been groundbreaking in their discovery of how mutations in Escherichia coli are either maintained or lost in the population (207). However, we still have much to learn about bacterial speciation, particularly in "wild" populations.

Freshwater lakes are ideal environments for studying bacterial speciation because they function like islands in the landscape. Several bacterial taxa can be found in freshwater globally, allowing cross-lake comparisons. Lakes have already been used as model systems to study evolution - loss of diversity in *Chlorobiales* of Trout Bog was used as an indicator of a strain with beneficial mutation quickly outcompeting all other strains (54), and diversity within populations from the same lake and from different lakes was assessed using single amplified genomes (SAGs) (186). In this second study, sequence discrete populations (having less than 95% average nucleotide identity) classified as the same tribe in the established freshwater taxonomy (32) (having greater than 98% identical 16S rRNA genes) were observed co-existing in the same lakes. Some of these groups even had highly correlated abundance over time. Presumably, these closely related populations can co-existed because they do not out-compete one another; that is, they occupy slightly different ecological niches.

Using the metatranscriptomic dataset introduced in Chapter 5, we analyze gene expression in closely related freshwater taxa to learn if differences in expression can reveal niche partitioning. A similar analysis was performed in marine systems and found differences in the expression of transporters (182); therefore, we focused on expressed transporters in this analysis.

Methods

For detailed information on metatranscriptomic sampling and sequencing, please see Chapter 5. The data presented in this chapter diverges at the mapping stage - instead of mapping to the nonredundant coding region database created from published and newly sequenced freshwater genomes and metagenomic assemblies, metatranscriptomic reads were mapped to our reference database of curated MAGs and SAGs used in prior studies. This mapping was performed at 95% identity to capture reads from sequence discrete populations. Following mapping, reads were counted using FeatureCounts and normalized to transcripts/L using the internal standard as previously described. Count tables were processed in R (R Core Team, 2018) and plots were created using ggplot2 (Wickham, 2009) and cowplot (Wilke, 2017). Average nucleotide identity (ANI) was calculated using the “pairwise ANI” tool in the Integrated Microbial Genomes (IMG) database (137). No abundance threshold was imposed in this analysis. Correlation in expression was measured using Pearson’s correlation.

Results

Within each lake, we identified clusters of closely related MAGs or SAGs with sufficient numbers of mapped reads to identify expressed transporters. Genomes in these clusters were typically classified as the same freshwater tribe and had average nucleotide identities of 75% to 99%. Overall expression of these groups was highly correlated, with r2 values ranging from 0.84 to 1. Because these genomes are incomplete, only expression of genes found in all genomes was considered.

LD12 in Sparkling Lake

In Sparkling Lake, we analyzed expression of the freshwater tribe LD12. LD12 is commonly known as "the freshwater SAR11," as this heterotrophic *Alphaproteobacterium* is related to this ubiquitous marine group and is also widespread in freshwater (208). LD12 has recently been cultured and assigned the name *Candidatus* Fonsibacter ubiquis (209). This taxon is proposed to degrade low molecular weight compounds such as carboxylic acids and amino acids, and to have auxotrophies for some vitamins and amino acids (92).

Table 6.1. Genomes used in this study. Metatranscriptomes sequenced for the analysis of day vs. night expression in Chapter 5 were mapped to our curated collection of MAGs and SAGs. Clusters of closely related, expressed genomes from each lake were identified and further analyzed. Genomes used in this analysis are reported here.

IMG Genome ID	Genome Name	Classification	Lake of origin	Lake expressed	Completeness
2236347069	D10	LD12	Mendota	Sparkling	82%
2236661000	L09	LD12	Sparkling	Sparkling	68%
2264265190	D14	acI-A7	Sparkling	Sparkling	48%
2236661005	N04	acI-A7	Damariscotta	Sparkling	80%
2582580642	TE567	PnecC	Trout Bog	Trout Bog	72%
2582580500	TH238	PnecC	Trout Bog	Trout Bog	66%
2582580711	TH944	PnecC	Trout Bog	Trout Bog	70%
2593339188	TE788	PnecC	Trout Bog	Trout Bog	57%

In Sparkling Lake, two LD12 genomes with 96% ANI were expressed (Table 6.1). Their expression over time was correlated at r2 = 0.96. Many transporters, such as those for C4 dicarboxylates, amino acids, and long-chain fatty acids, were expressed in both populations (Figure 6.1). However, the population mapping D10 expressed an ammonium transporter and an antimicrobial peptide transporter not expressed in the population mapping to L09. Conversely, transporters for mannitol and potassium were expressed in L09 but D10.

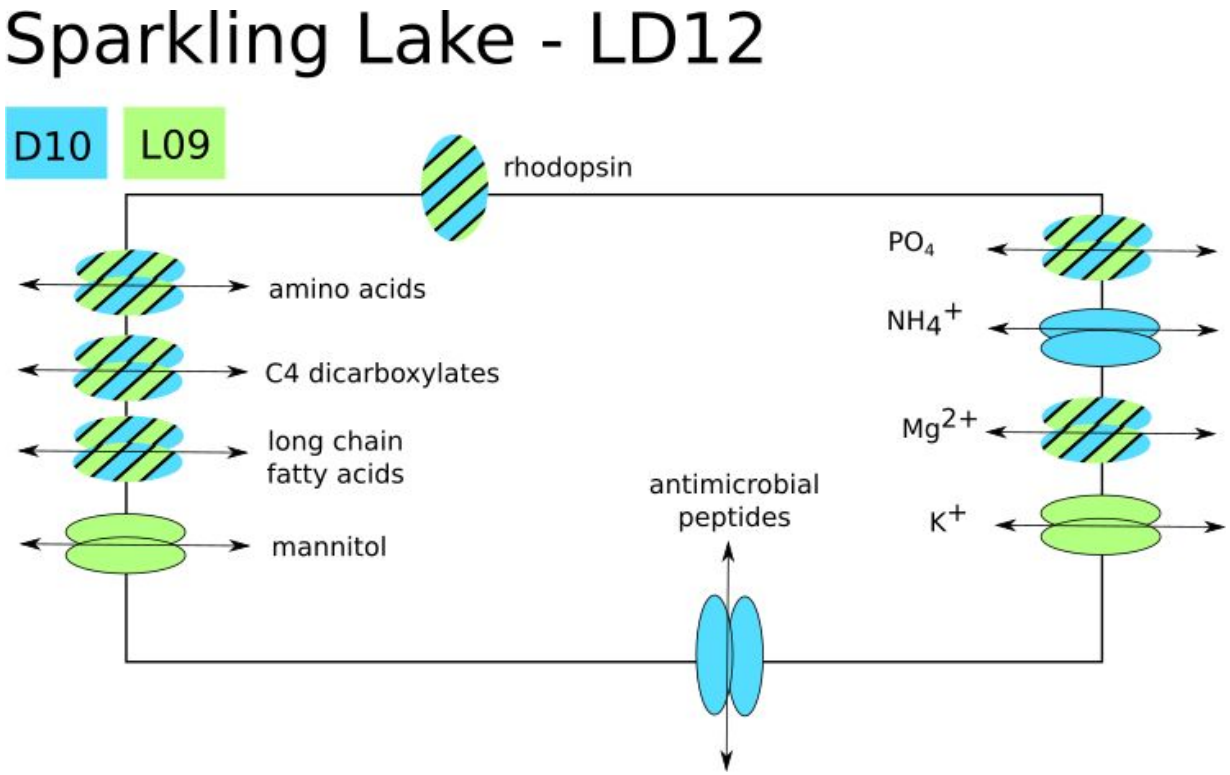

Figure 6.1. Expression of transporters in Sparkling Lake's LD12 populations. Different transporters suggest that niche partitioning in these populations occurs at the level of RNA regulation. The expression of overlapping carbon compound transporters may explain their correlated expression over time, while expression of different ion transporters, mannitol transport, and antimicrobial peptide transport may explain their co-existence.

acI-A7 in Sparkling Lake

Two other genomes, classified as acI-A7 with an ANI of 86%, were also co-expressed in Sparkling Lake with an r2 of 0.99 (Table 6.1, Figure 6.2). The ubiquitous freshwater clade acI has been proposed to consume amino acids, xylose, ribose, and polyamines (41, 117). These groups also possess highly streamlined genomes (34). Both populations expressed transporters for amino acids, multiple sugars, peptides,

ammonium, phosphate, polyamines, and xylose. Only one of the genomes, N04, expressed a transporter for melibiose, as well as a galactosidase that could potential break the disaccharide bond in melibiose. While no melibiose transporters were found in D14, several genes encoding galactosidases were found in both acI-A7 genomes. Transporters expressed in D14 but not N04 included potassium, polysaccharides, and vitamin B3 (nicotinamide mononucleotide).

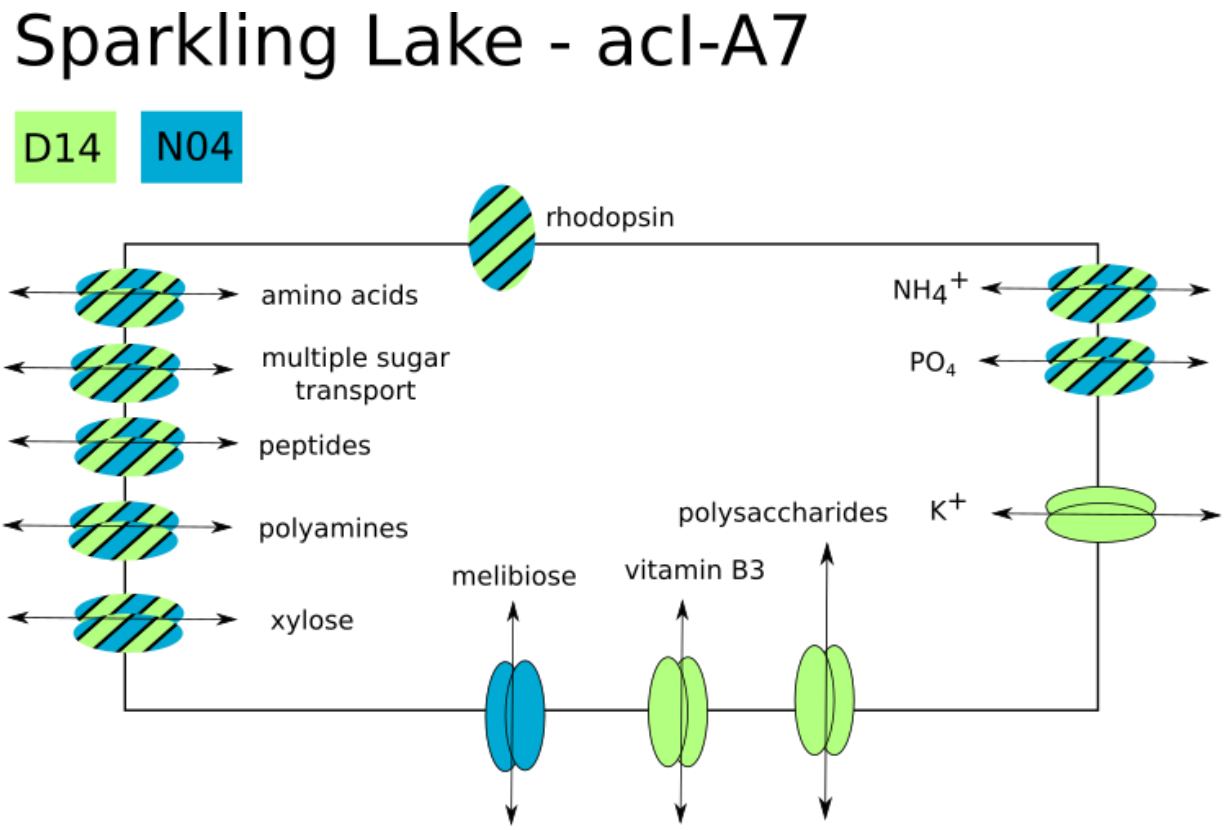

Figure 6.2. Expressed transporters in acI-A7 in Sparkling Lake. Two genetically similar populations of acI-A7 had many shared expressed transporters, but also interesting differences. These include transporters for vitamin B3 and polysaccharides.

PnecC in Trout Bog

The freshwater tribe PnecC includes well-studied and isolated members of *Polynucleobacter* such as *P. necessarius*. This species includes both subtypes that are obligate endosymbionts of freshwater ciliates and subtypes that are free-living (210). Free-living members of PnecC are thought to specialize in the degradation of low-weight compounds such as carboxylic acids, likely derived from photodegradation of dissolved organic carbon (36). High levels of genetic diversity were found within PnecC populations with nearly identical 16S regions (211).

We identified four expressed genomes in Trout Bog classified as PnecC. Two of these genomes, TH238 and TE567, had an ANI of 99%, indicating that they were likely recovered from the same population; these genomes were considered a single entity in the following analysis. ANI between the rest of the genomes was between 75-76%. These genomes ranged in completeness from 57% to 72%, and had strongly correlated expression (r2 = 1 for all pairs). Many transporters were shared between these populations, including those for dicarboxylates, tricarboxylates, amino acids, polyamines, biopolymers, and general sugars. Transporters for ions were also largely shared. However, there were several key differences. The population represented by TE567 and TH238 expressed a transporter for lipids, while the other two populations did not. TE788 did not express a tungstate transporter that was expressed in the other

two populations, and TH944 and TE788 expressed a multidrug resistance transporter not expressed in TE567 and TH238.

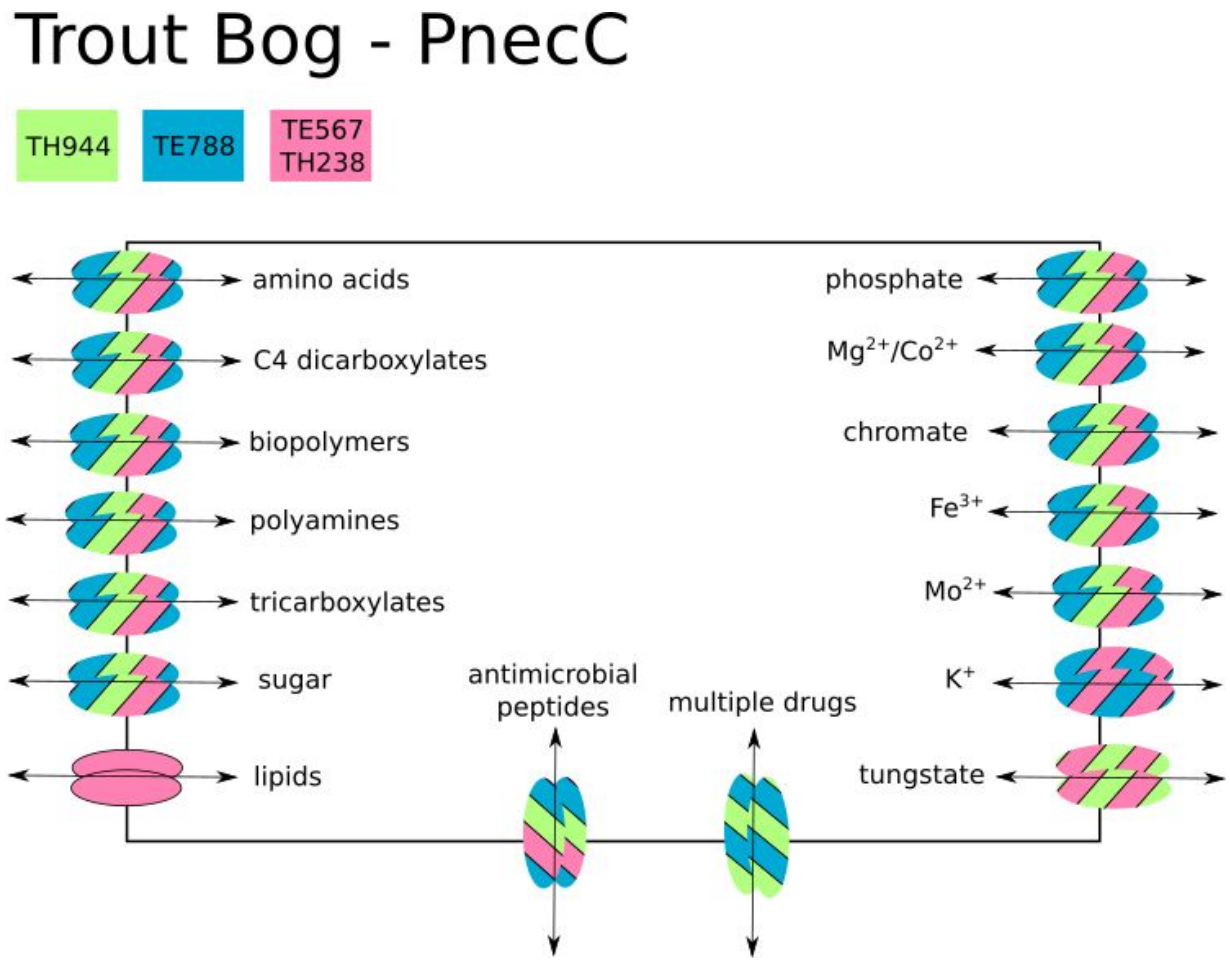

Figure 6.3. PnecC's expressed in transporters Trout Bog. Three genetically similar populations of PnecC were identified in Trout Bog. One was represented by two genomes, TE567 and TH238, with high ANI (99%). Many transporters were co-expressed in these populations, but key differences include transporters for tungstate, multidrug resistance, and lipids.

In Lake Mendota, clusters of expressed, genetically similar genomes classified as LD12 and acI were identified. However, no differential expression of transporters was found in these genomes.

Discussion

We hypothesized that expression of transporters may reveal how co-existing, genetically similar freshwater taxa fulfill different ecological niches. While not all of these populations displayed patterns in differential transporter expression, we did identify instance where this was the case. In particular, populations of LD12 and acI-A7 in Sparkling Lake and PnecC in Trout Bog had correlated gene expression over time, but different expressed transporters. The lack of this trend in LD12 and acI from Lake Mendota is potentially a function of the reference genomes used, as SAGs from LD12 and acI were particularly targeted in a sequencing project on Sparkling Lake, and MAGs of PnecC were particularly plentiful in Trout Bog.

Many of the transporters in these populations were expressed across each tribe, such as dicarboxylate and long-chain fatty acid transporters in LD12, sugar and polyamine transporters in acI-A7, and carboxylates, biopolymers, and polyamines in PnecC. This overlap may explain why total expression of these related organisms is so tightly correlated over time. Differentially expressed transporters included ammonium and antimicrobial peptide transporters in LD12, polysaccharide and vitamin B3 transporters in acI-A7, and lipid, tungstate, and multdrug transporters in PnecC. This suggests slightly different nutrient requirements in these populations during the time of

sampling, potentially enabling co-existence. The frequent appearance of expressed genes encoding antimicrobial peptide, multidrug, and antibiotic transporters in the populations presented here and in genomes not shown is intriguing, as chemical warfare has not been well-studied in freshwater. Transcription of these genes suggests that negative interactions may be an unaccounted factor structuring freshwater microbial communities.

Since closely related, co-existing bacteria were first observed, niche partitioning in these groups has been a mystery. Here, we present evidence suggesting that niche partitioning in these populations can occur via regulation of transcription. This mode of niche partitioning is inherently highly variable. Depending on the availability of nutrients and carbon substrates, these groups may be in direct competition or occupying completely different niches. Still, changes in transcription appear sufficient to create diverging populations. Our analysis of expression in closely related freshwater taxa provides insight into how bacterial speciation can occur in wild populations by modifying gene regulation.

Chapter 7: Perspectives and future research

Continuing the time series

Time series analysis is a powerful technique in the experimental design arsenal and one that is particularly useful for ecosystem testing, where rigorous positive and negative controls are often impossible to find. The time series approaches used in this thesis were effective at determining community composition and variability, improving bin quality of metagenome assemblies, and capturing transient gene expression. These experiments depended on the rich history of microbial and environmental data from our study sites, which informed my study design and provided context to my results. As such, one top priority for the McMahon Lab should be to continue collecting and analyzing samples for long-term time series.

Several projects that depend on the continuation of the time series are already in the works. For example, the metagenomic time series in Trout Bog and Lake Mendota have proved useful for studying virus-host dynamics in freshwater, and work is continuing on this project (131). The 16S rRNA gene amplicon time series is also well-suited as a test dataset for comparing different bioinformatics methods and determining best practices in data analysis. Continuing perturbations in Lake Mendota, such as the invasions of the zebra mussel and the spiny water flea (212), provide an unfortunate but timely opportunity to compare microbial communities before and after

food-web shifting events. Finally, given the pivotal role of freshwater (and particularly bogs) in global carbon cycling, maintaining the time series as the climate changes is a research area of high importance.

MAGs and SAGs as a reference database

In this work, I present an analysis of nearly 200 MAGs from multiple lakes and a diversity of taxonomic groups, as well as newly generated bins and SAGs from our study sites in conjunction with the metatranscriptomic study. While broad-scale genome analysis is useful for relating microbial metabolisms to ecosystem functioning, studies of a single taxonomic group or function can also be of great interest. Our genomes have proven to be of use to collaborators studying virus-host interactions, as they can quickly access the predicted metabolic functions of host bacteria. Another project that utilizes our genomes investigates nitrogen fixation in Trout Bog, specifically how different types of nitrogenases in diverse taxonomic groups are adapted to unique niches, while another analyzes rhodopsin genes to expand the known diversity of non-photosynthetic light harvesting in freshwater.

One observation in the MAGs that has spawned a larger research project is genes homologous to those encoding extracellular electron transfer (EET). These MAGs, recovered from Trout Bog, are from diverse taxonomic groups such as Chlorobi and Betaproteobacteria that have not previously been shown to perform EET. Still, the presence of EET seemed likely, given research demonstrating that humic substances, the main component of dissolved organic carbon in Trout Bog, can be used as an

electron carrier by EET organisms. If present, this metabolic process would be an unaccounted energy source in freshwater. To test for the presence of EET organisms, an undergraduate researcher and I built microbial fuel cells inoculated with water from Trout Bog and amended with acetate. These fuel cells quickly began to produce current, while a negative control inoculated with filtered Trout Bog water did not. A similar setup suspended from a buoy in Trout Bog also produced current, indicating that EET is occurring in Trout Bog. This project is likely to continue in the next few years.

Finally, these genomes can be paired with other types of data to improve our knowledge of metabolic functions. Programs such as PICRUST infer metabolic function from 16S sequences based on relatedness to cultured, sequenced isolates. However, these programs perform terribly on environmental data (confirmed by myself for freshwater), likely because the public collection of cultured isolates is biased towards members of the human microbiome. A more appropriate collection of references for our freshwater lakes may be genomes obtained from the same study sites, even if potential functions have not been confirmed experimentally. One can envision using the MAGs and SAGs from my projects to build a PICRUST-like program specifically for use with our long-term 16S time series.

Further insights from GEODES

Similar to the MAGs, I have chosen to present a only a handful of stories from my metatranscriptomic data, but this data also provides a wealth of information about the taxa and functions in freshwater microbial communities. Time series

metatranscriptomics, paired with comprehensive environmental datasets, can inform studies of people's favorite freshwater microbes. At a minimum, metatranscriptomics can confirm that an organism detected via DNA is alive and doing something, instead of existing in a dormant state or having lysed its DNA to the environment long before sampling. Within a single organism, studying co-expressed genes may reveal linked functions, regulatory mechanisms, or even new pathways that may not be detected using only DNA. Metatranscriptomics can be paired with other projects on these study sites to provide more information about target freshwater microbes.

This short-term metatranscriptomic study also lays the groundwork for metatranscriptomic work on longer time scales. Incorporating gene expression data into our decadal time series would likely reveal adaptations and community changes more quickly than metagenomics or 16S sequencing. However, expression changes so rapidly that interpreting a long-term time series of metatranscriptomics would be impossible without understanding the potential for short-term variability. My metatranscriptomics research revealed drastic changes in RNA profiles within four hours, and likely a finer resolution time series would have found changes on the order of minutes. While testing seasonal or annual trends in gene expression may be a worthwhile endeavor, the high variability of these data make any such investigation challenging.

Culturing

A major weakness of microbial ecology is the need to infer metabolic functions based on gene annotations. Currently, this is the best way to analyze uncultured

microbes on a large scale, but another alternative would be to expand our repertoire of cultured freshwater microbes (213). Even while I have been in graduate school, several "unculturable" ubiquitous freshwater bacteria have been isolated. Great strides have been made isolating and characterizing species of *Polynucleobacter*, which possesses such high levels of diversity within strains with nearly identical 16S rRNA genes that new species may be proposed (211, 214, 215). Isolates of *Limnohabitans,* a fast-growing, heavily grazed freshwater microbe, has undergone many new metabolic tests in culture (216, 217). The methylotrophic *Methylotenera* (including LD28 in the freshwater taxonomy) had previously been isolated from sediments in Lake Washington, but related members of Methylophilaceae have since been isolated from a freshwater water column (173). While not closely related to *Methylotenera,* these new isolates are still methylotrophs and now comprise a new genus, *Methylopumilis*. Another ubiquitous group, LD12, is known as the "freshwater SAR11" because it is the only freshwater-adapted clade of *Pelagibacteriales*. Like SAR11, LD12 has a streamlined genome and was resistant to isolations until recently. This research has revealed novel pathways for *Pelagibacteriales*, potential mechanisms of adaptation from saline to freshwater, and has led to the proposed species name *Candidatus* Fonsibacter ubiquis (209). Finally, the highly abundant freshwater clade acI had previously been enriched, but had not be cultured until catalase was added (42). This was a surprising result because genomic evidence suggested that acI produces its own catalase.

Clearly, culturing provides a much needed foundation to the hypothesis-generating sequencing methods in freshwater microbial ecology. I hope that the data presented in this thesis will guide future culturing efforts and that my hypotheses will one day be tested using freshwater isolates or co-cultures.

Conclusions

In this thesis, I use time series of sequencing data to explore how microbial metabolisms scale to ecosystem functions. This approach provided insight into how microbes function in communities. I learned that freshwater microbial communities are incredibly diverse and dynamic through the 16S rRNA gene amplicon time series of bog lakes. My analysis of MAGs and functional marker genes demonstrated that metabolic processes differ by lake, but not as much as I was expecting. Steps from different freshwater biogeochemical cycles were connected through multi-functional microbes, and taxonomy appeared to play a role in how carbon and nutrients were processed. By comparing gene expression in day versus night in freshwater, I discovered that primary production and sugar degradation occur at different times. This suggested that metabolites are exchanged between phototrophic and heterotrophic microbial community members, and that community organization in freshwater is analogous to plant cells. Finally, I compared expression in closely related freshwater taxa to propose mechanisms of niche partitioning based on transcriptional regulation. My thesis work has significantly advanced both our knowledge of microbial-mediated biogeochemical cycling in freshwater and our understanding of microbial community properties.

Bibliography

1. Nydahl A, Panigrahi S, Wikner J. 2013. Increased microbial activity in a warmer and wetter climate enhances the risk of coastal hypoxia. FEMS Microbiol Ecol 85:338–347.

2. Christensen GA, Somenahally AC, Moberly JG, Miller CM, King AJ, Gilmour CC, Brown SD, Podar M, Brandt CC, Brooks SC, Palumbo AV, Wall JD, Elias DA. 2018. Carbon Amendments Alter Microbial Community Structure and Net Mercury Methylation Potential in Sediments. Appl Environ Microbiol 84.

3. McClain ME, Boyer EW, Lisa Dent C, Gergel SE, Grimm NB, Groffman PM, Hart SC, Harvey JW, Johnston CA, Mayorga E, McDowell WH, Pinay G. 2003. Biogeochemical Hot Spots and Hot Moments at the Interface of Terrestrial and Aquatic Ecosystems. Ecosystems 6:301–312.

4. Williamson CE, Saros JE, Vincent WF, Smol JP. 2009. Lakes and reservoirs as sentinels, integrators, and regulators of climate change. Limnol Oceanogr 54:2273–2282.

5. Lauber CL, Ramirez KS, Aanderud Z, Lennon J, Fierer N. 2013. Temporal variability in soil microbial communities across land-use types. ISME J 7:1641–1650.

6. Gilbert JA, Steele JA, Caporaso JG, Steinbrück 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.

7. Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, Breitbart M, Buchanan J, Desnues C, Dinsdale E, Edwards R, Felts B, Haynes M, Liu H, Lipson D, Mahaffy J, Martin-Cuadrado AB, Mira A, Nulton J, Pasić L, Rayhawk S, Rodriguez-Mueller J, Rodriguez-Valera F, Salamon P, Srinagesh S, Thingstad TF, Tran T, Thurber RV, Willner D, Youle M, Rohwer F. 2010. Viral and microbial community dynamics in four aquatic environments. ISME J 4:739–751.

8. Cole JJ, Prairie YT, Caraco NF, McDowell WH, Tranvik LJ, Striegl RG, Duarte CM, Kortelainen P, Downing JA, Middelburg JJ, Melack J. 2007. Plumbing the Global Carbon Cycle: Integrating Inland Waters into the Terrestrial Carbon Budget. Ecosystems 10:172–185.

9. Houghton RA. 2007. Balancing the Global Carbon Budget. Annu Rev Earth Planet Sci 35:313–347.

10. Manning AC, Keeling RF. 2006. Global oceanic and land biotic carbon sinks from the Scripps atmospheric oxygen flask sampling network. Tellus B Chem Phys Meteorol 58.

11. Seitzinger S, Harrison JA, Böhlke JK, Bouwman AF, Lowrance R, Peterson B, Tobias C, Van Drecht G. 2006. Denitrification across landscapes and waterscapes: a synthesis. Ecol Appl

16:2064–2090.

12. Holgerson MA, Raymond PA. 2016. Large contribution to inland water CO2 and CH4 emissions from very small ponds. Nat Geosci 9:222–226.

13. Brune A, Frenzel P, Cypionka H. 2000. Life at the oxic-anoxic interface: microbial activities and adaptations. FEMS Microbiol Rev 24:691–710.

14. Logares R, Bråte J, Bertilsson S, Clasen JL, Shalchian-Tabrizi K, Rengefors K. 2009. Infrequent marine-freshwater transitions in the microbial world. Trends Microbiol 17:414–422.

15. Eiler A, Mondav R, Sinclair L, Fernandez-Vidal L, Scofield DG, Schwientek P, Martinez-Garcia M, Torrents D, McMahon KD, Andersson SG, Stepanauskas R, Woyke T, Bertilsson S. 2016. Tuning fresh: radiation through rewiring of central metabolism in streamlined bacteria. ISME J 10:1902–1914.

16. Clasen JL, Brigden SM, Payet JP, Suttle CA. 2008. Evidence that viral abundance across oceans and lakes is driven by different biological factors. Freshw Biol 53:1090–1100.

17. Pomeroy LR, Wiebe WJ. 1988. Energetics of microbial food webs. Hydrobiologia 159:7–18.

18. Pomeroy L, Williams PL, Azam F, Hobbie J. 2007. The Microbial Loop. Oceanography 20:28–33.

19. del Giorgio PA, Cole JJ, Cimbleris A. 1997. Respiration rates in bacteria exceed phytoplankton production in unproductive aquatic systems. Nature 385:148–151.

20. Cotner JB, Biddanda BA. 2002. Small Players, Large Role: Microbial Influence on Biogeochemical Processes in Pelagic Aquatic Ecosystems. Ecosystems 5:105–121.

21. Kritzberg ES, Cole JJ, Pace ML, Granéli W, Bade DL. 2004. Autochthonous versus allochthonous carbon sources of bacteria: Results from whole-lake 13C addition experiments. Limnol Oceanogr 49:588–596.

22. Jansson M, Bergstrom A-K, Blomqvist P, Drakare S. 2000. Allochthonous Organic Carbon and Phytoplankton/Bacterioplankton Production Relationships in Lakes. Ecology 81:3250.

23. Lebreton B, Pollack JB, Blomberg B, Palmer TA, Adams L, Guillou G, Montagna PA. 2016. Origin, composition and quality of suspended particulate organic matter in relation to freshwater inflow in a South Texas estuary. Estuar Coast Shelf Sci 170:70–82.

24. Allgaier M, Grossart HP. 2006. Seasonal dynamics and phylogenetic diversity of free-living and particle-associated bacterial communities in four lakes in northeastern Germany. Aquat Microb Ecol 45:115–128.

25. Selje N, Simon M. 2003. Composition and dynamics of particle-associated and free-living bacterial communities in the Weser estuary, Germany. Aquat Microb Ecol 30:221–237.

26. Lovley DR, Goodwin S. 1988. Hydrogen concentrations as an indicator of the predominant terminal electron-accepting reactions in aquatic sediments. Geochim Cosmochim Acta 52:2993–3003.

27. Capone DG, Kiene RP. 1988. Comparison of microbial dynamics in marine and freshwater sediments: Contrasts in anaerobic carbon catabolism. Limnol Oceanogr 33:725–749.

28. Lovley DR, Coates JD, Blunt-Harris EL, Phillips EJP, Woodward JC. 1996. Humic substances as electron acceptors for microbial respiration. Nature 382:445–448.

29. O'Malley MA. 2008. "Everything is everywhere: but the environment selects": ubiquitous distribution and ecological determinism in microbial biogeography. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 39:314–325.

30. Jones SE, McMahon KD. 2009. Species-sorting may explain an apparent minimal effect of immigration on freshwater bacterial community dynamics. Environ Microbiol 11:905–913.

31. Zwart G, Crump BC, Agterveld MPK, Hagen F, Han SK. 2002. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquat Microb Ecol 28:141–155.

32. 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.

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

34. Livermore JA, Emrich SJ, Tan J, Jones SE. 2014. Freshwater bacterial lifestyles inferred from comparative genomics. Environ Microbiol 16:746–758.

35. He S, Stevens SLR, Chan L-K, Bertilsson S, Glavina Del Rio T, Tringe SG, Malmstrom RR, McMahon KD. 2017. Ecophysiology of Freshwater Verrucomicrobia Inferred from Metagenome-Assembled Genomes. mSphere 2.

36. Hahn MW, Scheuerl T, Jezberová J, Koll U, Jezbera J, Šimek K, Vannini C, Petroni G, Wu QL. 2012. The passive yet successful way of planktonic life: genomic and experimental analysis of the ecology of a free-living polynucleobacter population. PLoS One 7:e32772.

37. Allgaier M, Grossart H-P. 2006. Diversity and seasonal dynamics of Actinobacteria populations in four lakes in northeastern Germany. Appl Environ Microbiol 72:3489–3497.

38. Kontur WS, Schackwitz WS, Ivanova N, Martin J, Labutti K, Deshpande S, Tice HN, Pennacchio C, Sodergren E, Weinstock GM, Noguera DR, Donohue TJ. 2012. Revised sequence and annotation of the *Rhodobacter sphaeroides* 2.4.1 genome. J Bacteriol 194:7016–7017.

39. Tank M, Liu Z, Frigaard N-U, Tomsho LP, Schuster SC, Bryant DA. 2017. Complete Genome Sequence of the Photoautotrophic and Bacteriochlorophyll e -Synthesizing Green Sulfur

Bacterium *Chlorobaculum limnaeum* DSM 1677 T. Genome Announc 5:e00529–17.

40. Hahn MW. 2003. Isolation of strains belonging to the cosmopolitan Polynucleobacter necessarius cluster from freshwater habitats located in three climatic zones. Appl Environ Microbiol 69:5248–5254.

41. Garcia SL, McMahon KD, Grossart H-P, Warnecke F. 2013. Successful enrichment of the ubiquitous freshwater acI Actinobacteria. Environ Microbiol Rep 6:21–27.

42. Kim S, Kang I, Seo J-H, Cho J-C. 2018. Culturing the ubiquitous freshwater actinobacterial acI lineage by supplying a biochemical "helper" catalase.

43. Prosser JI. 2010. Replicate or lie. Environ Microbiol 12:1806–1810.

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

45. Linz AM, Crary BC, Shade A, Owens S, Gilbert JA, Knight R, McMahon KD. 2017. Bacterial Community Composition and Dynamics Spanning Five Years in Freshwater Bog Lakes. mSphere 2.

46. Lennon JT. 2011. Replication, lies and lesser-known truths regarding experimental design in environmental microbiology. Environ Microbiol 13:1383–1386.

47. Magnuson JJ. 1990. Long-Term Ecological Research and the Invisible Present. Bioscience 40:495–501.

48. Shade A, Gregory Caporaso J, Handelsman J, Knight R, Fierer N. 2013. A meta-analysis of changes in bacterial and archaeal communities with time. ISME J 7:1493–1506.

49. Jones SE, Cadkin TA, Newton RJ, McMahon KD. 2012. Spatial and temporal scales of aquatic bacterial beta diversity. Front Microbiol 3.

50. 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.

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

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

53. Gifford SM, Sharma S, Moran MA. 2014. Linking activity and function to ecosystem dynamics in a coastal bacterioplankton community. Front Microbiol 5.

54. Bendall ML, Stevens SL, Chan L-K, Malfatti S, Schwientek P, Tremblay J, Schackwitz W,

Martin J, Pati A, Bushnell B, Froula J, Kang D, Tringe SG, Bertilsson S, Moran MA, Shade A, Newton RJ, McMahon KD, Malmstrom RR. 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J 10:1589–1601.

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

56. Yannarell AC, Kent AD, Lauster GH, Kratz TK, Triplett EW. 2003. Temporal patterns in bacterial communities in three temperate lakes of different trophic status. Microb Ecol 46:391–405.

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

58. Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, Xia LC, Xu ZZ, Ursell L, Alm EJ, Birmingham A, Cram JA, Fuhrman JA, Raes J, Sun F, Zhou J, Knight R. 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. ISME J 10:1669–1681.

59. Brooks JP, Edwards DJ, Harwich MD Jr, Rivera MC, Fettweis JM, Serrano MG, Reris RA, Sheth NU, Huang B, Girerd P, Vaginal Microbiome Consortium, Strauss JF 3rd, Jefferson KK, Buck GA. 2015. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. BMC Microbiol 15:66.

60. Tremblay J, Singh K, Fern A, Kirton ES, He S, Woyke T, Lee J, Chen F, Dangl JL, Tringe SG. 2015. Primer and platform effects on 16S rRNA tag sequencing. Front Microbiol 6:771.

61. Ramachandran A, Walsh DA. 2015. Investigation of XoxF methanol dehydrogenases reveals new methylotrophic bacteria in pelagic marine and freshwater ecosystems. FEMS Microbiol Ecol 91:fiv105.

62. Peura S, Sinclair L, Bertilsson S, Eiler A. 2015. Metagenomic insights into strategies of aerobic and anaerobic carbon and nitrogen transformation in boreal lakes. Sci Rep 5:12102.

63. Luef B, Frischkorn KR, Wrighton KC, Holman H-YN, Birarda G, Thomas BC, Singh A, Williams KH, Siegerist CE, Tringe SG, Downing KH, Comolli LR, Banfield JF. 2015. Diverse uncultivated ultra-small bacterial cells in groundwater. Nat Commun 6:6372.

64. Hamilton JJ, Garcia SL, Brown BS, Oyserman BO, Moya-Flores F, Bertilsson S, Malmstrom RR, Forest KT, McMahon KD. 2017. Metabolic Network Analysis and Metatranscriptomics Reveals Auxotrophies and Nutrient Sources of the Cosmopolitan Freshwater Microbial Lineage acI.

65. Moran MA, Satinsky B, Gifford SM, Luo H, Rivers A, Chan L-K, Meng J, Durham BP, Shen C, Varaljay VA, Smith CB, Yager PL, Hopkinson BM. 2013. Sizing up metatranscriptomics.

ISME J 7:237–243.

66. Tsementzi D, Poretsky R, Rodriguez-R LM, Luo C, Konstantinidis KT. 2014. Evaluation of metatranscriptomic protocols and application to the study of freshwater microbial communities. Environ Microbiol Rep 6:640–655.

67. Aylward FO, Eppley JM, Smith JM, Chavez FP, Scholin CA, DeLong EF. 2015. Microbial community transcriptional networks are conserved in three domains at ocean basin scales. Proc Natl Acad Sci U S A 112:5443–5448.

68. Bowers RM, Doud DFR, Woyke T. 2017. Analysis of single-cell genome sequences of bacteria and archaea. Emerg Top Life Sci 1:249–255.

69. Doud DFR, Woyke T. 2017. Novel approaches in function-driven single-cell genomics. FEMS Microbiol Rev 41:538–548.

70. Hedlund BP, Dodsworth JA, Murugapiran SK, Rinke C, Woyke T. 2014. Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter." Extremophiles 18:865–875.

71. Coyte KZ, Schluter J, Foster KR. 2015. The ecology of the microbiome: Networks, competition, and stability. Science 350:663–666.

72. Herren C, McMahon K. 2017. Cohesion: A method for quantifying the connectivity of microbial communities.

73. Pradeep Ram AS, Colombet J, Perriere F, Thouvenot A, Sime-Ngando T. 2015. Viral and grazer regulation of prokaryotic growth efficiency in temperate freshwater pelagic environments. FEMS Microbiol Ecol 91:1–12.

74. Grujcic V, Nuy JK, Salcher MM, Shabarova T, Kasalicky V, Boenigk J, Jensen M, Simek K. 2018. Cryptophyta as major bacterivores in freshwater summer plankton. ISME J.

75. Salcher MM, Ewert C, Šimek K, Kasalický V, Posch T. 2016. Interspecific competition and protistan grazing affect the coexistence of freshwater Betaproteobacterial strains. FEMS Microbiol Ecol 92.

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

77. Paver SF, Kent AD. 2010. Temporal Patterns in Glycolate-Utilizing Bacterial Community Composition Correlate with Phytoplankton Population Dynamics in Humic Lakes. Microb Ecol 60:406–418.

78. Simek K, Kasalický V, Zapomělová E, Hornák K. 2011. Alga-derived substrates select for distinct Betaproteobacterial lineages and contribute to niche separation in Limnohabitans strains. Appl Environ Microbiol 77:7307–7315.

79. Aharonovich D, Sher D. 2016. Transcriptional response of Prochlorococcus to co-culture

with a marine Alteromonas: differences between strains and the involvement of putative infochemicals. ISME J 10:2892–2906.

80. Schatz D, Rosenwasser S, Malitsky S, Wolf SG, Feldmesser E, Vardi A. 2017. Communication via extracellular vesicles enhances viral infection of a cosmopolitan alga. Nat Microbiol 2:1485–1492.

81. Lovley DR. 2017. Happy together: microbial communities that hook up to swap electrons. ISME J 11:327–336.

82. Magnuson JJ, Kratz TK, Benson BJ. 2006. Long-term Dynamics of Lakes in the Landscape: Long-term Ecological Research on North Temperate Lakes. Oxford University Press on Demand.

83. Taipale S, Jones RI, 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.

84. Garcia SL, Salka I, Grossart H-P, Warnecke F. 2013. Depth-discrete profiles of bacterial communities reveal pronounced spatio-temporal dynamics related to lake stratification. Environ Microbiol Rep 5:549–555.

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

86. Eiler A, Heinrich F, Bertilsson S. 2011. Coherent dynamics and association networks among lake bacterioplankton taxa. ISME J 6:330–342.

87. 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.

88. Brock TD. 2012. A Eutrophic Lake: Lake Mendota, Wisconsin. Springer Science & Business Media.

89. Kara EL, Hanson PC, Hu YH, Winslow L, McMahon KD. 2012. A decade of seasonal dynamics and co-occurrences within freshwater bacterioplankton communities from eutrophic Lake Mendota, WI, USA. ISME J 7:680–684.

90. Hall MW, Rohwer RR, Perrie J, McMahon KD, Beiko RG. 2017. Ananke: temporal clustering reveals ecological dynamics of microbial communities. PeerJ 5:e3812.

91. Martinez-Garcia M, Swan BK, Poulton NJ, Gomez ML, Masland D, Sieracki ME, Stepanauskas R. 2012. High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. ISME J 6:113–123.

92. Ghylin TW, Garcia SL, Moya F, Oyserman BO, Schwientek P, Forest KT, Mutschler J,

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

93. Gilbert J. 2012. The Earth Microbiome Project: A new paradigm in geospatial and temporal studies of microbial ecology. SciVee.

94. 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.

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

96. 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.

97. Caporaso JG, Gregory Caporaso J, 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.

98. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez A, Knight R. 2017. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems 2:e00191–16.

99. 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.

100. 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.

101. McGinnis S, Madden TL. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32:W20–W25.

102. 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.

103. Rohwer RR, Hamilton JJ, Newton RJ, McMahon KD. 2017. TaxAss: Leveraging Custom

Databases Achieves Fine-Scale Taxonomic Resolution.

104. McMurdie PJ, Holmes S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One 8:e61217.

105. Jafari SM, Zarre S, Alavipanah SK, Ghahremaninejad F. 2013. Exploring the generality of associations between plant functional traits: evidence within ecological groups along an altitudinal gradient in Hyrcanian forest. Plant Species Biol 29:E31–E39.

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

107. Newton RJ, Shade A. 2016. Lifestyles of rarity: understanding heterotrophic strategies to inform the ecology of the microbial rare biosphere. Aquat Microb Ecol 78:51–63.

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

109. Kalyuzhnaya MG, Beck DAC, Vorobev A, Smalley N, Kunkel DD, Lidstrom ME, Chistoserdova L. 2012. Novel methylotrophic isolates from lake sediment, description of Methylotenera versatilis sp. nov. and emended description of the genus Methylotenera. Int J Syst Evol Microbiol 62:106–111.

110. Herren CM, Webert KC, McMahon KD. 2016. Environmental Disturbances Decrease the Variability of Microbial Populations within Periphyton. mSystems 1.

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

112. Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S. 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proc Natl Acad Sci U S A 103:13104–13109.

113. Cram JA, Chow C-ET, Sachdeva R, Needham DM, Parada AE, Steele JA, Fuhrman JA. 2015. Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. ISME J 9:563–580.

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

115. Rusak JA, Jones SE, Kent AD, Shade AL, McMahon TM. 2009. Spatial synchrony in microbial community dynamics: testing among-year and lake patterns. SIL Proceedings, 1922-2010 30:936–940.

116. Ivars-Martinez E, Martin-Cuadrado A-B, D'Auria G, Mira A, Ferriera S, Johnson J, Friedman R, Rodriguez-Valera F. 2008. Comparative genomics of two ecotypes of the marine planktonic copiotroph Alteromonas macleodii suggests alternative lifestyles

associated with different kinds of particulate organic matter. ISME J 2:1194–1212.

117. Hamilton JJ, Garcia SL, Brown BS, Oyserman BO, Moya-Flores F, Bertilsson S, Malmstrom RR, Forest KT, McMahon KD. 2017. Metabolic Network Analysis and Metatranscriptomics Reveal Auxotrophies and Nutrient Sources of the Cosmopolitan Freshwater Microbial Lineage acI. mSystems 2.

118. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048.

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

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

121. Roberts I, Houlden A. 2013. Faculty of 1000 evaluation for A meta-analysis of changes in bacterial and archaeal communities with time. F1000 - Post-publication peer review of the biomedical literature.

122. Wasmund K, Ipek KurtbÃ¶ke D, Burns KA, Bourne DG. 2009. Microbial diversity in sediments associated with a shallow methane seep in the tropical Timor Sea of Australia reveals a novel aerobic methanotroph diversity. FEMS Microbiol Ecol 68:142–151.

123. Williamson CE, Dodds W, Kratz TK, Palmer MA. 2008. Lakes and streams as sentinels of environmental change in terrestrial and atmospheric processes. Front Ecol Environ 6:247–254.

124. Butman D, Stackpoole S, Stets E, McDonald CP, Clow DW, Striegl RG. 2016. Aquatic carbon cycling in the conterminous United States and implications for terrestrial carbon accounting. Proc Natl Acad Sci U S A 113:58–63.

125. Simek K, Kasalicky V, Jezbera J, Jezberova J, Hahn MW. 2010. Broad Habitat Range of the Phylogenetically Narrow R-BT065 Cluster, Representing a Core Group of the Betaproteobacterial Genus Limnohabitans. Appl Environ Microbiol 76:3763–3763.

126. Salcher MM, Posch T, Pernthaler J. 2012. In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. ISME J 7:896–907.

127. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963.

128. Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H, Hugenholtz P. 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME J 4:642–647.

129. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open

source tool for metagenomics.

130. Rohwer RR, Hamilton JJ, Newton RJ, McMahon KD. 2017. TaxAss: Leveraging a Custom Freshwater Database Achieves Fine-Scale Taxonomic Resolution.

131. Roux S, Chan L-K, Egan R, Malmstrom RR, McMahon KD, Sullivan MB. 2017. Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. Nat Commun 8:858.

132. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18.

133. Sommer DD, Delcher AL, Salzberg SL, Pop M. 2007. Minimus: a fast, lightweight genome assembler. BMC Bioinformatics 8:64.

134. Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J. 2012. Ray Meta: scalable de novo metagenome assembly and profiling. Genome Biol 13:R122.

135. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165.

136. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595.

137. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. 2012. IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res 40:D115–22.

138. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431–437.

139. Darling AE, Jospin G, Lowe E, Matsen FA 4th, Bik HM, Eisen JA. 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ 2:e243.

140. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS, Banfield JF. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun 7:13219.

141. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

142. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST : architecture and applications. BMC Bioinformatics 10:421.

143. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods 9:811–814.

144. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Genome Standards Consortium, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731.

145. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445–51.

146. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

147. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A. 2015. Microbial species delineation using whole genome sequences. Nucleic Acids Res 43:6761–6771.

148. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One 5:e9490.

149. Hong S, Bunge J, Leslin C, Jeon S, Epstein SS. 2009. Polymerase chain reaction primers miss half of rRNA microbial diversity. ISME J 3:1365–1373.

150. Beversdorf LJ, Miller TR, McMahon KD. 2013. The Role of Nitrogen Fixation in Cyanobacterial Bloom Toxicity in a Temperate, Eutrophic Lake. PLoS One 8:e56103.

151. Remsen CC, Carpenter EJ, Schroeder BW. 1972. Competition for Urea among Estuarine Microorganisms. Ecology 53:921–926.

152. Jorgensen N. 1998. Effects of sunlight on occurrence and bacterial turnover of specific carbon and nitrogen compounds in lake water. FEMS Microbiol Ecol 25:217–227.

153. Berman T, Bronk DA. 2003. Dissolved organic nitrogen: a dynamic participant in aquatic ecosystems. Aquat Microb Ecol 31:279–305.

154. Mou X, Vila-Costa M, Sun S, Zhao W, Sharma S, Moran MA. 2011. Metatranscriptomic

signature of exogenous polyamine utilization by coastal bacterioplankton. Environ Microbiol Rep 3:798–806.

155. Igarashi K, Kashiwagi K. 1999. Polyamine transport in bacteria and yeast. Biochem J 344:633.

156. Bulushi IA, Al Bulushi I, Poole S, Deeth HC, Dykes GA. 2009. Biogenic Amines in Fish: Roles in Intoxication, Spoilage, and Nitrosamine Formation—A Review. Crit Rev Food Sci Nutr 49:369–377.

157. Acquisti C, Kumar S, Elser JJ. 2009. Signatures of nitrogen limitation in the elemental composition of the proteins involved in the metabolic apparatus. Proc Biol Sci 276:2605–2610.

158. Bragg JG, Wagner A. 2009. Protein material costs: single atoms can make an evolutionary difference. Trends Genet 25:5–8.

159. Holkenbrink C, Barbas SO, Mellerup A, Otaki H, -U. Frigaard N. 2011. Sulfur globule oxidation in green sulfur bacteria is dependent on the dissimilatory sulfite reductase system. Microbiology 157:1229–1239.

160. Bowles MW, Mogollón JM, Kasten S, Zabel M, Hinrichs K-U. 2014. Global rates of marine sulfate reduction and implications for sub-sea-floor metabolic activities. Science 344:889–891.

161. Karhunen J, Arvola L, Peura S, Tiirola M. 2013. Green sulphur bacteria as a component of the photosynthetic plankton community in small dimictic humic lakes with an anoxic hypolimnion. Aquat Microb Ecol 68:267–272.

162. Kanao T, Kawamura M, Fukui T, Atomi H, Imanaka T. 2002. Characterization of isocitrate dehydrogenase from the green sulfur bacterium Chlorobium limicola. A carbon dioxide-fixing enzyme in the reductive tricarboxylic acid cycle. Eur J Biochem 269:1926–1931.

163. Tang K-H, Blankenship RE. 2010. Both forward and reverse TCA cycles operate in green sulfur bacteria. J Biol Chem 285:35848–35854.

164. Hanson TE, Tabita FR. 2001. A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from *Chlorobium tepidum* that is involved with sulfur metabolism and the response to oxidative stress. Proc Natl Acad Sci U S A 98:4397–4402.

165. Bertilsson S, Tranvik LJ. 1998. Photochemically produced carboxylic acids as substrates for freshwater bacterioplankton. Limnol Oceanogr 43:885–895.

166. Ramanan R, Kim B-H, Cho D-H, Oh H-M, Kim H-S. 2016. Algae–bacteria interactions: Evolution, ecology and emerging applications. Biotechnol Adv 34:14–29.

167. Gong X, Garcia-Robledo E, Lund MB, Lehner P, Borisov SM, Klimant I, Revsbech NP, Schramm A. 2018. Gene expression of terminal oxidases in two marine bacterial strains

exposed to nanomolar oxygen concentrations. FEMS Microbiol Ecol.

168. Peters JW, Schut GJ, Boyd ES, Mulder DW, Shepard EM, Broderick JB, King PW, Adams MWW. 2015. [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1853:1350–1369.

169. Giroldo D, Vieira AAH. 2005. Polymeric and free sugars released by three phytoplanktonic species from a freshwater tropical eutrophic reservoir. J Plankton Res 27:695–705.

170. Boden R, Hutt LP, Rae AW. 2017. Reclassification of Thiobacillus aquaesulis (Wood & Kelly, 1995) as *Annwoodia aquaesulis* gen. nov., comb. nov., transfer of Thiobacillus (Beijerinck, 1904) from the Hydrogenophilales to the Nitrosomonadales, proposal of Hydrogenophilalia class. nov. within the "Proteobacteria", and four new families within the orders Nitrosomonadales and Rhodocyclales. Int J Syst Evol Microbiol 67:1191–1205.

171. Bowman JP, Sly LI, Stackebrandt E. 1995. The phylogenetic position of the family Methylococcaceae. Int J Syst Bacteriol 45:182–185.

172. Latypova E, Yang S, Wang Y-S, Wang T, Chavkin TA, Hackett M, Schäfer H, Kalyuzhnaya MG. 2010. Genetics of the glutamate-mediated methylamine utilization pathway in the facultative methylotrophic beta-proteobacterium Methyloversatilis universalis FAM5. Mol Microbiol 75:426–439.

173. Salcher MM, Neuenschwander SM, Posch T, Pernthaler J. 2015. The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. ISME J 9:2442–2453.

174. Chistoserdova L, Kalyuzhnaya MG, Lidstrom ME. 2009. The expanding world of methylotrophic metabolism. Annu Rev Microbiol 63:477–499.

175. Beversdorf LJ, Chaston SD, Miller TR, McMahon KD. 2015. Microcystin mcyA and mcyE Gene Abundances Are Not Appropriate Indicators of Microcystin Concentrations in Lakes. PLoS One 10:e0125353.

176. Walsh JR, Munoz SE, Jake Vander Zanden M. 2016. Outbreak of an undetected invasive species triggered by a climate anomaly. Ecosphere 7:e01628.

177. van Gemerden H, Montesinos E, Mas J, Guerrero R. 1985. Diel cycle of metabolism of phototrophic purple sulfur bacteria in Lake Cisó (Spain). Limnol Oceanogr 30:932–943.

178. Staehr PA, Bade D, Van de Bogert MC, Koch GR, Williamson C, Hanson P, Cole JJ, Kratz T. 2010. Lake metabolism and the diel oxygen technique: State of the science. Limnol Oceanogr Methods 8:628–644.

179. Horne AJ. 1975. Algal nitrogen fixation in Californian streams: diel cycles and nocturnal fixation. Freshw Biol 5:471–477.

180. Vila-Costa M, Sharma S, Moran MA, Casamayor EO. 2013. Diel gene expression profiles

of a phosphorus limited mountain lake using metatranscriptomics. Environ Microbiol 15:1190–1203.

181. Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA. 2009. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. Environ Microbiol 11:1358–1375.

182. Ottesen EA, Young CR, Eppley JM, Ryan JP, Chavez FP, Scholin CA, DeLong EF. 2013. Pattern and synchrony of gene expression among sympatric marine microbial populations. Proc Natl Acad Sci U S A 110:E488–97.

183. Cole JJ. 1982. Interactions Between Bacteria and Algae in Aquatic Ecosystems. Annu Rev Ecol Syst 13:291–314.

184. Paver SF, Hayek KR, Gano KA, Fagen JR, Brown CT, Davis-Richardson AG, Crabb DB, Rosario-Passapera R, Giongo A, Triplett EW, Kent AD. 2013. Interactions between specific phytoplankton and bacteria affect lake bacterial community succession. Environ Microbiol 15:2489–2504.

185. Goldstone JV, Pullin MJ, Bertilsson S, Voelker BM. 2002. Reactions of hydroxyl radical with humic substances: bleaching, mineralization, and production of bioavailable carbon substrates. Environ Sci Technol 36:364–372.

186. Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T, Tringe SG, Andersson SGE, Bertilsson S, Malmstrom RR, McMahon KD. 2018. Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. ISME J 12:742–755.

187. Satinsky BM, Gifford SM, Crump BC, Moran MA. 2013. Use of internal standards for quantitative metatranscriptome and metagenome analysis. Methods Enzymol 531:237–250.

188. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, Le Chatelier E, Pelletier E, Bonde I, Nielsen T, Manichanh C, Arumugam M, Batto J-M, Quintanilha Dos Santos MB, Blom N, Borruel N, Burgdorf KS, Boumezbeur F, Casellas F, Doré J, Dworzynski P, Guarner F, Hansen T, Hildebrand F, Kaas RS, Kennedy S, Kristiansen K, Kultima JR, Léonard P, Levenez F, Lund O, Moumen B, Le Paslier D, Pons N, Pedersen O, Prifti E, Qin J, Raes J, Sørensen S, Tap J, Tims S, Ussery DW, Yamada T, MetaHIT Consortium, Renault P, Sicheritz-Ponten T, Bork P, Wang J, Brunak S, Ehrlich SD, MetaHIT Consortium. 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat Biotechnol 32:822–828.

189. Bushnell B. 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner.

190. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its

applications to single-cell sequencing. J Comput Biol 19:455–477.

191. Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217.

192. Pruitt KD. 2001. RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Res 29:137–140.

193. Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26:680–682.

194. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930.

195. Yang E-W, Girke T, Jiang T. 2013. Differential gene expression analysis using coexpression and RNA-Seq data. Bioinformatics 29:2153–2161.

196. Gupta RS, Lorenzini E. 2007. Phylogeny and molecular signatures (conserved proteins and indels) that are specific for the Bacteroidetes and Chlorobi species. BMC Evol Biol 7:71.

197. Bertilsson S, Tranvik LJ. 2000. Photochemical transformation of dissolved organic matter in lakes. Limnol Oceanogr 45:753–762.

198. Graf A, Schlereth A, Stitt M, Smith AM. 2010. Circadian control of carbohydrate availability for growth in Arabidopsis plants at night. Proc Natl Acad Sci U S A 107:9458–9463.

199. Vitova M, Bisova K, Kawano S, Zachleder V. 2015. Accumulation of energy reserves in algae: From cell cycles to biotechnological applications. Biotechnol Adv 33:1204–1218.

200. Kind T, Meissen JK, Yang D, Nocito F, Vaniya A, Cheng Y-S, Vandergheynst JS, Fiehn O. 2012. Qualitative analysis of algal secretions with multiple mass spectrometric platforms. J Chromatogr A 1244:139–147.

201. Hellebust JA. 1965. Excretion of some organic compounds by marine phytoplankton. Limnol Oceanogr 10:192–206.

202. Bauld J, Brock TD. 1974. Algal excretion and bacterial assimilation in hot spring algal mats. J Phycol 10:101–106.

203. Sharp JH. 1977. Excretion of organic matter by marine phytoplankton: Do healthy cells do it? Limnol Oceanogr 22:381–399.

204. Adolph S, Bach S, Blondel M, Cueff A, Moreau M, Pohnert G, Poulet SA, Wichard T, Zuccaro A. 2004. Cytotoxicity of diatom-derived oxylipins in organisms belonging to different phyla. J Exp Biol 207:2935–2946.

205. Braakman R, Follows MJ, Chisholm SW. 2017. Metabolic evolution and the

self-organization of ecosystems. Proc Natl Acad Sci U S A 114:E3091–E3100.

206. Nelson CE, Goldberg SJ, Wegley Kelly L, Haas AF, Smith JE, Rohwer F, Carlson CA. 2013. Coral and macroalgal exudates vary in neutral sugar composition and differentially enrich reef bacterioplankton lineages. ISME J 7:962–979.

207. Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular evolution over 60,000 generations. Nature 551:45–50.

208. Paver S, Muratore DJ, Newton RJ, Coleman M. 2018. Re-evaluating the salty divide: phylogenetic specificity of transitions between marine and freshwater systems.

209. Henson MW, Lanclos VC, Faircloth BC, Thrash JC. 2018. Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. ISME J.

210. Hahn MW, Lang E, Brandt U, Wu QL, Scheuerl T. 2009. Emended description of the genus Polynucleobacter and the species *Polynucleobacter necessarius* and proposal of two subspecies, *P. necessarius* subsp. *necessarius* subsp. nov. and *P. necessarius* subsp. *asymbioticus* subsp. nov. Int J Syst Evol Microbiol 59:2002–2009.

211. Hahn MW, Jezberová J, Koll U, Saueressig-Beck T, Schmidt J. 2016. Complete ecological isolation and cryptic diversity in Polynucleobacter bacteria not resolved by 16S rRNA gene sequences. ISME J 10:1642–1655.

212. Walsh JR, Carpenter SR, Vander Zanden MJ. 2016. Invasive species triggers a massive loss of ecosystem services through a trophic cascade. Proc Natl Acad Sci U S A 113:4081–4085.

213. Salcher MM, Šimek K. 2016. Isolation and cultivation of planktonic freshwater microbes is essential for a comprehensive understanding of their ecology. Aquat Microb Ecol 77:183–196.

214. Hahn MW, Koll U, Jezberová J, Camacho A. 2015. Global phylogeography of pelagic Polynucleobacter bacteria: restricted geographic distribution of subgroups, isolation by distance and influence of climate. Environ Microbiol 17:829–840.

215. Hoetzinger M, Schmidt J, Jezberová J, Koll U, Hahn MW. 2017. Microdiversification of a Pelagic Polynucleobacter Species Is Mainly Driven by Acquisition of Genomic Islands from a Partially Interspecific Gene Pool. Appl Environ Microbiol 83.

216. Horňák K, Kasalický V, Šimek K, Grossart H-P. 2017. Strain-specific consumption and transformation of alga-derived dissolved organic matter by members of the *Limnohabitans*-C and *Polynucleobacter*-B clusters of Betaproteobacteria. Environ Microbiol 19:4519–4535.

217. Kasalický V, Jezbera J, Hahn MW, Šimek K. 2013. The diversity of the *Limnohabitans* genus, an important group of freshwater bacterioplankton, by characterization of 35 isolated strains. PLoS One 8:e58209.