# Metagenome-assembled genomes reveal metabolic similarities and differences between freshwater lakes

Alexandra M. Linz, Shaomei He, Samuel Schmitz, Sarah Stevens, Rex R. Malmstrom, Robin R. Rohwer, Joshua J. Hamilton, Karthik Anantharaman, Tijana Glavina del Rio, Stefan Bertilsson, Katherine D. McMahon

# Abstract

Bacterial communities play a major role in freshwater nutrient cycling, yet relatively little is known about the individuals in these communities. Freshwater microbial communities are diverse and dynamic, and most microbes cannot be cultivated in the laboratory, making the study of specific taxa difficult. Identifying which bacterial taxa facilitate specific biogeochemical transformations would improve our understanding of freshwater nutrient cycling. In this research, we use time-series resolved metagenome-assembled genomes (MAGs) to link bacterial taxa to their functions in two contrasting freshwater lakes. Prediction of metabolic pathways based on gene content in these genomes is used to infer metabolic traits of wild bacterial populations. The two lakes chosen for this research have different chemical characteristics; Trout Bog is a humic lake, while Lake Mendota is eutrophic. Therefore, we expected to find significant differences in gene content between our study sites. We found several key differences between Trout Bog and Lake Mendota in carbon fixation pathways, degradation of complex carbon compounds, and nitrogen usage. However, we also observed striking similarities, such as an overlap in sugar degradation pathways, enrichment for genes encoding the same steps of the sulfur cycle, high numbers of genes related to polyamine metabolism, and the presence of methylotrophy and anoxygenic aerobic photosynthesis in both lakes. While further experiments are needed to confirm metabolic traits in these populations of bacteria, this work lays a foundation by beginning to pull apart the complexity of freshwater bacterial communities and their role in nutrient cycling.

# Introduction

Freshwater lakes are important to nutrient cycling on a global scale. Lakes act as integrators of their surrounding landscapes, collecting nutrients from terrestrial ecosystems (Williamson et al. 2008). This makes lakes “hotspots” in the landscape, particularly in carbon cycling (Butman et al. 2015). Within lakes, much of this nutrient processing is performed by the microbial community. At the ecosystem level, the microbial community is often considered to be a single, unchanging entity, but previous research has revealed high levels of diversity and change over time in these communities (Kara et al. 2013; Linz et al. 2017). We seek to integrate genomic information about individual bacterial taxa into our understanding of freshwater bacterial communities to better explain how microbes contribute to ecosystem-level nutrient cycling.

Nutrient cycling in lakes has been previously thought of in terms of primary production vs. respiration, allochthonous vs. autochthonous carbon, dissolved vs particulate matter, and organic vs. inorganic nutrients (McGowan et al. 2016). These broad categories are frequently used to calculate budgets for lakes. However, the microbes responsible for most nutrient cycling in lakes are diverse in both taxonomy and function. One of the grand challenges in microbial ecology is to link taxonomic groups to ecosystem functions to improve predictive models and budgets of lake nutrient cycling. To reach this level of understanding, the broad categories of nutrient cycling must be broken down into microbe-specific categories.

Because the majority of freshwater microbes cannot yet be cultured, sequencing data can be used instead to infer the function of specific taxonomic groups. Metagenomics has previously been used to shed light into the role of aquatic microbes. For example, this type of study has been used to investigate functional differences between salt and freshwater microbes (Eiler et al. 2015), to identify important functions in humic lakes (Peura et al. 2012), and to compare microbial communities on a global scale (Gimmler et al. 2016). However, many metagenomics-based studies take a gene-centric approach to investigating microbial functions. In this study, we use metagenome-assembled genomes (MAGs) to compare microbial functions between two lakes of different trophic statuses with an organism-centric approach instead. Analyzing genomes rather than genes provides better insight into the ecological roles of specific microbes within freshwater communities.

Previously, we used time series metagenomics to assemble nearly 200 high-quality MAGs from Lake Mendota, a highly productive eutrophic lake, and Trout Bog, a humic bog lake. Genomes from this dataset have been used to study genome sweeps in Trout Bog (Bendall et al. 2016), to build metabolic networks of the ubiquitous freshwater Actinobacteria acI in Lake Mendota (Hamilton et al. 2017), and to propose functions for freshwater Verrucomicrobia (He et al. 2017). Lake Mendota and Trout Bog were chosen as the study sites for 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>), NSF, Emily Stanley, Center for Limnology, University of Wisconsin-Madison), their previous 16S time series analyses (Hall et al. 2017; Linz et al. 2017), and because of their contrasting chemical limnology (Table 1). We hypothesized that we would be able to infer information about ecosystem-level functions based on genomic content in our comprehensive dataset from a eutrophic and a humic lake. This includes differences in primary production between lakes, preferences for degradation of simple or complex carbon compounds, and biases towards certain steps in the nitrogen and sulfur cycles based on the availability of these inorganic compounds. Our analysis of these MAGs demonstrated many similarities in microbial functioning across freshwater ecosystems, but also revealed key differences.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lake Mendota | Trout Bog Epilimnion | Trout Bog Hypolimnion |
| Location | Madison, WI | Boulder Junction, WI | |
| Depth of lake (m) | 25.3 | 7.9 | |
| Surface area of lake (km2) | 39.61 | 0.01 | |
| Microbial sampling depth range (m) | 0-12 | 0-2 | 2-7 |
| Oxygenation | Aerobic | Aerobic | Anaerobic |
| pH | 8.60 (0.35) | 5.00 (0.19) | 5.19 (0.24) |
| Dissolved inorganic carbon (ppm) | 40.68 (6.67) | 1.79 (1.52) | 6.16 (5.51) |
| Dissolved organic carbon (ppm) | 5.49 (0.92) | 18.10 (2.80) | 24.20 (5.51) |
| Total dissolved nitrogen (ppb) | 1070.38 (421.01) | 612.14 (153.12) | 1448.99 (1127.77) |
| Total nitrogen (ppb) | 1262.25 (353.04) | 754.45 (229.14) | 1711.86 (1509.75) |
| Total dissolved phosphorus (ppb) | 88.56 (57.53) | 13.45 (7.63) | 78.14 (95.81) |
| Total phosphorus (ppb) | 111.94 (47.11) | 27.12 (16.34) | 107.67 (122.66) |
| Sulfate (ppm) | 18.04 (3.16) | 1.22 (0.34) | 0.84 (0.60) |

**Table 1. Characteristics of Lake Mendota and Trout Bog.** The epilimnion of Lake Mendota and both layers of Trout Bog were sampled using an integrated water column for microbial DNA weekly during the ice-free periods in 2005, 2007, 2008, and 2009. Chemistry data was measured 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 (2005 – 2009), with standard deviations reported in parentheses.

# Methods

Samples were collected from Lake Mendota and Trout Bog as described in Bendall et al. 2016(Bendall et al. 2016). Briefly, integrated samples of the water column were collected during the ice-free periods of 2007-2009 in Trout Bog and 2005-2009 in Lake Mendota. In Lake Mendota, the top 12 meters of the water column were sampled, approximating the epilimnion. The epilimnion and hypolimnion 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 consistently 0-2 meters for the epilimnion and 2-7 meters for the hypolimnion. DNA was collected by filtering 150-250mL 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).

As described in Bendall et al. 2016, metagenomic sequencing was performed by the Department of Energy Joint Genome Institute (DOE JGI) (Walnut Creek, CA, USA). Samples were sequenced on the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA), except for six libraries (two from each lake-layer) sequenced using the Illumina TruSeq protocol on the Illumina GA IIx platform (Illumina). Metadata on the samples sequenced is available in <supp table>. Metagenomic reads were pooled and then assembled by lake and layer using SOAPdenovo (Luo et al. 2012) and Minimus (Sommer et al. 2007), resulting in assemblies for Lake Mendota, the Trout Bog epilimnion, and the Trout Bog hypolimnion containing information from multiple years of samples. Contigs from the combined assemblies were binned into metagenome assembled genomes (MAGs) using MetaBat (Kang et al. 2015) and metagenomic reads were mapped to the assembled contigs using the Burrows-Wheeler Aligner (Li and Durbin 2010), allowing time-series resolved binning. DOE JGI’s Integrated Microbial Genome (IMG) database tool (Markowitz et al. 2012) was used for gene annotation and prediction. MAG completeness was estimated based on the presence of a core set of genes used CheckM and Rinke et al. 2013 (Parks et al. 2015; Rinke et al. 2013) . The MAGs were classified using Phylosift (Darling et al. 2014).

Only MAGs that were at least 50% complete and must have passed quality checks for contamination using CheckM (Parks et al. 2015) were included in this study. A total of 194 high quality bacterial metagenome assembled genomes (MAGs) were recovered from a metagenomic time series in Trout Bog and Lake Mendota, ranging in completeness from 50-99%. Of the 194 MAGs, 100 were recovered from Lake Mendota, 31 were recovered from the epilimnion of Trout Bog, and 63 were recovered from the hypolimnion of Trout Bog <Table S1, MAG\_information.csv>. Several MAGs in the epilimnion and hypolimnion of Trout Bog appeared to be from the same population based on average nucleotide identities great than 99% using JGI’s ANI calculator (Varghese et al. 2015) <Table S2, ANI\_matrix.csv>. This is possible because genomes from the epilimnion and hypolimnion were assembled separately.

Pathways were analyzed by exporting IMG’s functional data on the MAGs, which included KEGG, COG, pfam, and TIGR annotations (cite?), and aggregating annotations by the pathways in which they participate. To be “present,” a pathway must have had 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 a unique step. Glycoside hydrolases were annotated using dbCAN (Yin et al. 2012). The tree presented in Figure 4 is an approximate maximum likelihood tree constructed in FastTree (Price, Dehal, and Arkin 2010) using whole genome alignments, as in <https://github.com/McMahonLab/Scripts/tree/master/Phylogeny>. As it has not been bootstrapped, it is not intended to infer evolutionary history, merely similarity between genomes. Tree building was performed in BioLinux 8 (Field et al. 2006). Marker gene analysis of sulfur cycling was performed using IMG’s functional search. Data formatting and plotting was performed in R (Team 2015) using the following packages: ggplot2 (Wickham 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 (Wickham 2007), and APE (Paradis, Claude, and Strimmer 2004). The datasets, scripts, and intermediate files used to predict pathway presence and absence are available at < github link>.

# Results/Discussion

## Overview of Dataset

The phylogenetic distribution of MAGs was consistent with the classifications of 16S rRNA gene amplicon sequencing results, presumably because MAGs were recovered from abundant populations in the community (Figure 1). These results are consistent with other 16S-based studies in these sites (Hall et al. 2017; Linz et al. 2017). We inferred metabolic potential of these abundant microbes based on the gene content in the recovered MAGs, focusing on carbon, nitrogen, and sulfur metabolisms.

<Fig 1a Barchart of RPKM of MAGs from metagenome mapping>

<Fig 1b Barchart of 16S phyla>

**Figure 1. How representative are the MAGs of the microbial community?** The taxonomic classifications of MAGs and their proportions of reads mapped from the metagenomic time series (A) reflect the community composition observed via 16S rRNA gene amplicon sequencing of the same samples (B).

## Primary Production and Phototrophy

Primary production is a critical component of the carbon cycle in lakes. Therefore, we looked at potential routes of primary production within the microbial community, expecting to find differences between our two ecosystems. In Lake Mendota, MAGs classified as Cyanobacteria comprised most photoautotrophs in the dataset. These populations contained genes encoding enzymes in the Calvin-Benson-Bassham (CBB) pathway. In Trout Bog, genomes appearing to be from photoautotrophic organisms were classified as *Chlorobium clathratiforme*, a species of Chlorobiales widespread in humic lakes (Karhunen et al. 2013). 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 (Kanao et al. 2002; Tang and Blankenship 2010). Both photoautotrophs contained genes potentially encoding nitrogen fixation as well. Genes annotated as ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO), the key enzyme in the CBB pathway, were observed in some of the Chlorobiales MAGs. The reductive TCA cycle is the only carbon fixation pathway known to be active in this taxon, and a RubisCo-like gene found in isolates of *Chlorobium* was associated with sulfur metabolism and oxidative stress (Hanson and Tabita 2001). Given this information, it seems likely that this gene encodes a function other than carbon fixation in the Chlorobiales MAGs. One major difference besides carbon fixation pathway between Cyanobacteria and Chlorobi is their oxygen requirements; Cyanobacteria are aerobic, while Chlorobiales are anaerobic. These two primary producers seem to perform similar ecosystem functions in their respective lakes, but oxygen availability drives both the type of microbe acting in this role and the pathways that it uses for primary production.

The potential for photoheterotrophy via the anoxygenic aerobic phototrophic pathway (Martinez-Garcia et al. 2012) was identified in several MAGs from both lakes. Proteobacteria, particularly some unclassified MAGs of Burkholderiales, most often contained the marker genes for this process, although these genes were not broadly shared across the phylum. A MAG of Acidobacteria from the Trout Bog epilimnion also contained genes suggesting anoxygenic aerobic phototrophy. Another form of photoheterotrophy previously identified in freshwater is the use of light-activated proteins such as rhodopsins. We observed genes encoding rhodopsins in MAGs from both lakes, but more frequently in MAGs from Lake Mendota classified as Actinobacteria and Bacteroidetes. MAGs from Trout Bog, especially the epilimnion, harbored much less diversity and a lower abundance of genes encoding rhodopsins than those from Lake Mendota. A likely explanation for this observation is that Lake Mendota is clearer than the humic-stained Trout Bog, leading to greater light availability in Lake Mendota. (cite Shaomei’s preprint when it comes out).

## Central Metabolism and Simple Carbon Degradation

Freshwater contains a variety of low-complexity carbon sources such as carbohydrates, carboxylic acids, and one-carbon compounds. While carbon in freshwater is often divided into autochthonous (originating within the lake) and allochthonous (derived from the surrounding landscape) carbon, this distinction is less clear for bacteria. For example, there is substantial overlap in algal exudates and cellulose breakdown products, and while one-carbon compounds such as methane are produced in the lake, they are often produced via the decomposition of allochthonous carbon. Therefore, we found it more informative to categorize the carbon degradation pathways observed in our dataset by carbon complexity (Fig. 2).

Central metabolism is often the entry point for the least complex carbon compounds, and central metabolic pathways may reveal how a bacterium is using a carbon compound. The TCA cycle, arguably the most central pathway in bacteria, was notably absent in MAGs classified as Tenericutes in Lake Mendota and in unclassified MAGs (potentially members of the candidate phyla radiation) in the hypolimnion of Trout Bog. This is consistent with previous research on Tenericutes and members of the candidate phyla radiation (Brown et al. 2015; Miles 1992). Genes encoding enzymes in the glyoxylate cycle, a variant of the TCA cycle that is used to produce biosynthetic intermediates when glucose is not available, were observed in Chlamydiae in Lake Mendota, Acidobacteria in Trout Bog, and in some Proteobacteria in both lakes. The pentose phosphate pathway, both oxidative and non-oxidative phases, was found in MAGs from most phyla.

As oxidative phosphorylation is an important part of central metabolism for aerobic bacteria, we investigated which types of cytochromes were encoded in our MAGs (Fig 2). Cytochrome c oxidases, both aa3- and cbb3-type, were widespread in both lakes and frequently observed in the same genomes. As aa3-type cytochromes are associated with high oxygen concentrations and cbb3-type cytochromes are associated with low oxygen concentrations, the presence of both types suggests the flexibility to operate under a range of oxygen concentrations. Of the quinol-based cytochromes, genes encoding cytochrome d were most often observed in MAGs from the hypolimnion of Trout Bog, while cytochrome aa3-600 was found only in MAGs classified as Bacteroidetes and Proteobacteria in the Trout Bog epilimnion and cytochrome o was observed only in a Chlamydia MAG from Lake Mendota. Alternative complex III was identified in MAGs of Verromicrobia in both lakes, in Acidobacteria in Trout Bog (both layers), and in Bacteroidetes and Planctomycetes in Lake Mendota. The presence and absence of cytochrome types by lake may reflect both the difference in oxygen concentrations and in bacterial community composition.

Similarly, hydrogen metabolism is an aspect of central metabolism that can influence other aspects of a microbe’s nutrient usage. Iron-only hydrogenases were found primarily in Trout Bog’s hypolimnion (Fig. 2), consistent with their previously identified presence in fermenting anaerobic bacteria (Peters et al. 2015). Genes encoding [Ni-Fe] hydrogenases of groups 1 and 2, involved in hydrogen uptake, sensing, and nitrogen fixation, were widespread in the hypolimnion of Trout Bog, found only in Chlorobiales MAGs in the epilimnion of Trout Bog, and more rarely observed in Lake Mendota. Group 3 [Ni-Fe] hydrogenases were identified in MAGs belonging to Cyanobacteria and Chlorobiales, consistent with the proposed function of Group 3d, which is to remove excess electrons produced by photosynthesis. Finally, the Group 4 reversible hydrogenase were observed primarily in the hypolimnion of Trout Bog. These different types of hydrogenases, particularly the greater prevalence of hydrogenases in the anoxic region of Trout Bog, point to more lake-specific adaptations.

Algae in freshwater have been documenting producing high molecular weight carbohydrates such as glucose, fucose, rhamnose, arabinose, galactose, mannose, and xylose (Giroldo, Augusto, and Vieira 2005). To identify linkages between algae and heterotrophic bacteria, we analyzed putative sugar degradation pathways in our MAGs. Genes encoding the pathway for mannose degradation appeared frequently in both lakes. Mannose feeds into glycolysis, and can be used as the sole source of carbon and energy in bacteria such as *Escherichia coli*; this may explain why it was observed so frequently. 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 in Lake Mendota, and members of Bacteroidetes, Ignavibacteria, and Verrucomicrobia in Trout Bog). Putative pathways for the degradation of galactose were often observed in these same MAGs. Xylose is a freshwater sugar which has already been identified as potential carbon source for streamlined Actinobacteria (Ghylin et al. 2014); this was confirmed in our MAGs, with Bacteroidetes, Planctomycetes, and Verrucomicrobia in Lake Mendota and Bacteroidetes and Verrucomicrobia in Trout Bog as additional potential xylose degraders. Genes for the degradation of glycolate, an acid produced by algae and consumed by heterotrophic bacteria (Paver et al. 2017), were identified in Cyanobacteria and Proteobacteria in Lake Mendota and in Acidobacteria, Proteobacteria, and Verrucomicrobia in Trout Bog.

Methylotrophy, the ability to grow solely on one carbon compounds such as methane or methanol, appears to be a likely metabolism in MAGs from both Trout Bog and Lake Mendota. Putative pathways for methanol degradation were found in MAGs classified as Methylophilales, while MAGs from Methylococcales were potential methane degraders. The MAGs of Methylophilales also likely degrade methylamines, based on the presence of genes encoding the N-methylglutamate pathway or the tetrahydrofolate pathway (Latypova et al. 2010; Salcher et al. 2015a). Methylotrophy in cultured freshwater isolates from these taxa is well-documented (Kalyuzhnaya et al. 2012; Salcher et al. 2015b); however, genes encoding methanol degradation were also identified in MAGs from taxa not typically known as methylotrophs. These included MAGs classified as Burkholderiales, Rhizobiales, and Nitrosomonadales in Trout Bog. Given the rapid rate at which the known diversity of methylotrophs is increasing, this finding is intriguing, but not surprising (Chistoserdova, Kalyuzhnaya, and Lidstrom 2009).

A close up of text on a white background

Description generated with high confidence**Fig 2. Metabolisms in Lake Mendota vs Trout Bog.** Carbon cycling between the two lakes was largely similar, with a few key differences. Carbon fixation is carried out by different taxa using different pathways. Dissimilatory sulfate reduction was more common in Trout Bog than in Lake Mendota. Degradation and biosynthesis of polyamines was prevalent in MAGs from both lakes. Rhodopsins were most often observed in MAGs of Actinobacteria and Bacteroidetes from Lake Mendota. Central metabolism showed partitioning by lake and phylum, likely reflecting differences in oxygen concentrations and evolutionary history.

## Complex Carbon Degradation

Biopolymers in freshwater can be autochthonous (ex. algal polysaccharides) or allochthonous (ex. cellulose). While degradation of these high-complexity carbon sources may require specialized enzymes, their wide availability and high yield of sugars make the ability to degrade complex carbon sources an advantageous trait. One way to analyze the ability to degrade high-complexity carbon is through genes annotated as glycoside hydrolases (GHs), enzymes that breakdown glycosidic bonds 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 autochthonous and allochthonous carbon sources (He et al. 2017). We expanded this analysis of glycoside hydrolases to the entire dataset to identify differences in complex carbon degradation between lakes.

The coding density of glycoside hydrolases – the percentage of coding regions in a MAG annotated as a glycoside hydrolase – immediately revealed differences between Trout Bog and Lake Mendota, and even between the epilimnion and hypolimnion of Trout Bog (Fig. 3). The MAGs with the highest coding densities were found in members of Bacteroidales, Ignavibacteriales, Sphingobacteriales, and Verrucomicrobiales in the Trout Bog hypolimnion. The last two of those orders also contained MAGs with glycoside hydrolases in Lake Mendota and the Trout Bog epilimnion, but the others did not. There were several orders with glycoside hydrolases unique to Lake Mendota, including Mycoplasmatales (Tenericutes), Cytophagales (Bacteroidetes), Planctomycetales (Planctomycetes), and Puniceicoccales (Verrucomicrobia). In accordance with their ability to breakdown biopolymers to sugars, these MAGs from both lakes also contain putative degradation pathways for a variety of sugars (Fig. 2). The diversity of glycoside hydrolases, an indicator of the number of substrates an organism can degrade, was correlated with glycoside hydrolase coding density (r2= 0.39, p < 0.01).

Several glycoside hydrolase families were abundant in Lake Mendota and in both layers of Trout Bog. Starting with the most abundant, these included GH109, GH74, and GH23. While the most abundant glycoside hydrolase genes were similar between lakes, the increased diversity of these genes in Trout Bog’s hypolimnion suggested differences between their profiles of glycoside hydrolases and therefore differences in the diversity and complexity of their carbon sources. Lake Mendota contained unique glycoside hydrolases belonging to the family GH13, which contain enzymes related to cellulose degradation. The only unique glycoside hydrolase in the Trout Bog epilimnion was GH62. The hypolimnion contained many more unique enzymes than Lake Mendota or the epilimnion of Trout Bog, the most abundant of which were GH129 and GH89, GH43\_12, GH44, GH66, and GH67.

The increased coding density and diversity of glycoside hydrolase genes in the Trout Bog hypolimnion suggest that the bacterial community in this region relies more on complex carbon sources than simple carbon sources compared to Lake Mendota or even the epilimnion of Trout Bog. This may be because primary production results in increased availability of low complexity carbon compounds in epilimnia, while terrestrially-derived complex carbon polymers may be a more important source of carbon in the hypolimnion of Trout Bog. However, the taxonomic profile of MAGs containing glycoside hydrolases differed by lake and layer, even when the profiles of glycoside hydrolases themselves were more similar.

A screenshot of a cell phone

Description generated with high confidence

**Figure 3. Glycoside hydrolase coding density.** Annotations of glycoside hydrolases were used as an indication of complex carbon degradation. Glycoside hydrolase coding density was calcuclated for each MAG and averaged by order and lake. 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 density.

## Nitrogen Cycling

Nitrogen availability is an important factor structuring freshwater bacterial communities. Bog lakes such as Trout Bog are generally considered to be nitrogen-limited ecosystem, and nitrogen is present in humic lakes is often bound in complex carbon compounds. Lake Mendota, as a eutrophic ecosystem, is considered to have excess nitrogen due to urban and agricultural pollution. However, this nitrogen is quickly incorporated into biomass and is not always readily available to microbes. Because of these different nitrogen regimes, we expected to see differences in the nitrogen metabolisms of Trout Bog vs Lake Mendota.

Properties of genomes themselves may provide information about nitrogen limitation (Bragg 2011). We observed a bias in MAGs from Trout Bog towards encoding amino acids with less nitrogen compared to MAGs from Lake Mendota using a Wilcoxon rank sum test (p = 0.02). GC content and estimated genome size, other potential indicators of nitrogen limitation often correlated with amino acid bias, were not significantly different between lakes (p = 0.78 and p = 0.16, respectively). While amino acid bias suggests that limiting nitrogen concentrations in Trout Bog may lead to selection for organisms encoding nitrogen-poor proteins, other factors may be more important in determining properties such as genome size and GC content.

Putative pathways related to nitrogen metabolism appeared at similar frequencies from MAGs in both lakes. Dissimilatory and assimilatory nitrate reduction appear to be present in similar numbers of genomes, and denitrification appeared slightly less often in genomes from both lakes. Urea degradation was predicted in MAGs of both lakes, consistent with research where urea was found to be a significant nitrogen source for freshwater bacteria, particularly in epilimnia (Jorgenson et al. 1998), and where algae and bacteria were observed to compete for urea in an estuarine system (Remsen, Carpenter, and Schroeder 1972). Genes encoding the biosynthesis and degradation of polyamines such as spermidine and putrescine, potentially important compounds in the freshwater dissolved organic nitrogen pool, were prevalent in MAGs from both lakes.

We expected nitrogen fixation to be more prevalent in genomes from Trout Bog, but found instead that similar numbers of genomes contain genes encoding the potential pathway in Lake Mendota as well. However, taxonomy revealed differences between the two ecosystems. In Lake Mendota, nitrogen fixation appears restricted to Cyanobacteria, Betaproteobacteria, and Gammaproteobacteria. This may be a factor in the documented links between cyanobacterial bloom toxicity and nitrogen fixation in Lake Mendota (Beversdorf, Miller, and McMahon 2013). MAGs containing genes encoding nitrogen fixation are more phylogenetically diverse in Trout Bog, including several classes of Proteobacteria, Acidobacteria, Verrucomicrobia, Chlorobi, and Bacteroidetes (Fig 2, Fig 4C). This result may also suggest adaptation to nitrogen limitation, as this trait is maintained more frequently in populations from Trout Bog. Interestingly, two MAGs of Chlorobiales (potentially from the same population based on ANI) contained genes encoding Fe-only nitrogenase in addition to Mo nitrogenase. Methane production has been observed as an inherent property of the Fe-only nitrogenases (Zheng et al. 2018), suggesting that Chlorobiales may be an additional source of methane in Trout Bog.

Genes potentially encoding the biosynthesis, degradation, and transport of the polyamines and non-proteinogenic amino acids such as putrescine, spermidine, and canavanine were widespread in both lakes. While there is some evidence for the importance of polyamines in aquatic systems (Mou et al. 2011), these compounds have been less studied in freshwater and their ecological role is not yet resolved. Our results lend support to the hypothesis that these compounds are important parts of the dissolved organic nitrogen pool in freshwater. Polyamines play a critical but poorly understood role in bacterial metabolism (Igarashi and Kashiwagi 1999), and the exchange of these nitrogen compounds may be a factor structuring freshwater bacterial communities. One likely source of polyamines is higher trophic levels such as fish or zooplankton, as these compounds can result from the decomposition of amino acids.

Although we expected to find major differences in nitrogen metabolisms between lakes, we instead found striking similarities. Despite the chemical differences between lakes, similar nitrogen compounds are likely consumed by similar numbers of taxa. However, the identity of those taxa is the key difference between taxa, with nitrogen fixation as an example of a pathway with variable phylogenetic diversity between our two study sites. While Lake Mendota, as a eutrophic lake, is generally considered to have an excess of nitrogen, nitrogen can temporarily become limiting during periods of high biomass (Beversdorf, Miller, and McMahon 2013). This may explain why populations of Cyanobacteria in Lake Mendota have the ability to fix nitrogen.

A close up of a map

Description generated with high confidence

**Fig 4. Nitrogen and sulfur cycling.** Proportions of MAGs containing nitrogen (A) and sulfur pathways are relatively similar between lakes. Nitrogen fixation (B) was restricted to Cyanobacteria and Proteobacteria in Lake Mendota, but was more phylogenetically diverse in Trout Bog. The tree was constructed with FastTree using whole genome alignments and is intended to show similarity between MAGs, not to imply evolutionary history.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Process | Marker Gene | Lake Mendota | Trout Bog Epilimnion | Trout Bog Hypolimnion |
| Sulfide oxidation | Sulfide:quinone reductase | 20% | 31% | 36% |
| Sulfide oxidation | Flavocytochrome c | 3% | 14% | 5% |
| Sulfite reduction or oxidation | Dissimilatory sulfite reductase | 0% | 6 % | 8% |
| Sulfur oxidation | sox gene cluster | 4% | 20% | 17% |
| Sulfite oxidation | Sulfate adenylyltransferase | 45% | 52% | 49% |
| Sulfite oxidation | Phosphoadenosine phosphosulfate reductase | 23% | 23% | 29% |
| Assimilatory sulfate reduction | Assimilatory sulfate reductase | 51% | 40% | 59% |

**Table 2. Marker genes for sulfur cycling pathways.** This table reports the percentage of MAGs from each site containing a coding region annotated as a gene considered indicative a step in the sulfur cycle. Genes indicating oxidation dominated over those indicating reduction.

## Sulfur Cycling

Sulfur is another element structuring freshwater bacterial communities. We found that, as with carbon and nitrogen cycling, the identity and MAGs encoding steps in the sulfur cycle were similar between Lake Mendota and Trout Bog (Fig. 2). To further investigate sulfur metabolism, we searched for sulfur-related marker genes in our MAGs. The marker genes chosen included flavocytochrome c (sulfide oxidation), dissimilatory sulfate reductase (sulfate reduction, sulfate oxidation in some phototrophs and other bacteria), thiosulfate dehydrogenase (SOX oxidation), and sulfate adenylyltransferase (sulfite oxidation). We also searched for anaerobic sulfite reductases but did not detect any genes with this annotation. Like the results of the sulfur pathway analysis, oxidation was observed more frequently than reduction (Table 2). Dissimilatory sulfite reductase was observed only in MAGs from Trout Bog, especially MAGs classified as Chlorobiales. Because this enzyme in known to operate in reverse in phototrophs such as Chlorobiales, this may indicate additional oxidation rather than the presence of a reductive sulfur pathway. As expected, assimilatory sulfate reduction was the most common pathway identified (Table 2). For sulfur use as a terminal electron acceptor or donor, oxidation pathways were more prevalent than reduction pathways in both systems (Fig. 4A). This suggests that in these lakes, sulfate is more commonly used in biosynthesis pathways, while reduced forms of sulfur are used as electron donors for energy pathways.

## MAGs over Time

Because these MAGs were recovered from metagenomic time series, we can assess trends in the MAGs over time using reads mapped from the time series to the MAGs as a proxy for abundance. Combining this data with metabolic information about each MAG provides further insight into their ecology. For example, the model methlyotrophs Methylococcales and Methylophilales show distinct trends over time in each study site (Fig. 5). Methylococcales are aerobes that consume methane; in lakes, oxygen is available in the epilimnion during stratification and throughout the water column during spring and fall mixing only, while methane is produced in the anaerobic hypolimnion and sediments. Therefore, Methylococcales can likely only thrive where both oxygen and methane are present. As expected, MAGs classified as Methylococcales are found in the Trout Bog hypolimnion only during fall mixing, a time when those conditions are met in the deeper regions of the lake. Methylococcales in the epilimnion of Trout Bog decrease in the spring and early summer and increase again in the fall, likely reflecting methane concentrations, as oxygen is more constant. In the epilimnion of Lake Mendota, which has less stable stratification than Trout Bog, Methylococcales periodically increases in the summer fall. One possible explanation for this trend could be the small scale, wind-driven mixes that occur in Lake Mendota. In contrast, Methylophilales was detected in both epilimnia and the Trout Bog hypolimnion, suggesting it may be a facultative anaerobe. The methylotrophic metabolism of this taxon, which likely utilized methanol and methylamines, does not require oxygen. Its abundance is relatively stable over time, suggesting that its niche and carbon sources are also stable in our A screenshot of a cell phone

Description generated with high confidencestudy sites.

## Figure 5. Methylotrophs over time. Reads mapping to our MAG dataset from their originating metagenomic time series was used as a proxy for abundance over time. Read counts were normalized to reads per kilobase per million mapped reads (RPKM) and aggregated by year to reveal seasonal trends. MAGs classified as the methylotrophs Methylococcales (A-C) and Methylophilales (D-F) are presented here. These taxa show different seasonal trends in different lakes and layers, likely reflecting their ecological niches.

## Conclusions

Our analysis of gene content in MAGs from Trout Bog and Lake Mendota revealed both striking similarities and strong contrasts. Interesting metabolisms found in both lakes include methylotrophy, aerobic anoxygenic photosynthesis, and polyamine degradation. There was substantial overlap in sugar degradation pathways found in MAGs from our two study sites, and an analysis of sulfur cycling pathways showed a preference towards using sulfate in biosynthesis and reduced sulfur compounds in energy generation in both lakes. Many of the differences we observed between lakes were not in the pathways themselves, but in the taxa with genes encoding those pathways – primary production and nitrogen fixation appear to be performed by distinct taxonomic groups. The pathways used for carbon fixation differ by lake, as do the types of hydrogenases found in our MAGS. Dissolved oxygen levels appear to be an important factor in determining which pathways, enzymes, or taxa are found in Lake Mendota vs. Trout Bog. However, dissolved organic carbon also plays a role – while sugar degradation pathways were similar across lakes, the density and diversity of glycoside hydrolases was much greater in Trout Bog than in Lake Mendota, suggesting an increased emphasis on the degradation of complex carbon sources such as humic substances.

Using our dataset of time series resolved MAGs, additional work with metatranscriptomic data, chemical limnology, or cultured isolates can now target specific processes and taxa identified as particularly interesting or important in freshwater. These analyses can also be used to compare other freshwater ecosystems to our two study sites, which will help us understand what metabolic processes are shared between all lakes. While genomic data suggests rather than proves the presence or importance of metabolic pathways in ecosystems, this research lays the groundwork to further our understanding of freshwater nutrient cycling.

# Acknowledgments

# References

Bendall, Matthew L et al. 2016. “Genome-Wide Selective Sweeps and Gene-Specific Sweeps in Natural Bacterial Populations.” *ISME journal* 10: 1589–1601. http://dx.doi.org/10.1038/ismej.2015.241.

Beversdorf, Lucas J, Todd R Miller, and Katherine D McMahon. 2013. “The Role of Nitrogen Fixation in Cyanobacterial Bloom Toxicity in a Temperate , Eutrophic Lake.” *PloS one* 8(2): 1–11.

Bragg, Jason G. 2011. “How Prochlorococcus Bacteria Use Nitrogen Sparingly in Their Proteins.” *Molecular Ecology* 20(1): 27–28.

Brown, Christopher T. et al. 2015. “Unusual Biology across a Group Comprising More than 15% of Domain Bacteria.” *Nature* 523(7559): 208–11.

Butman, David et al. 2015. “Aquatic Carbon Cycling in the Conterminous United States and Implications for Terrestrial Carbon Accounting.” *Proceedings of the National Academy of Sciences*: 1–6.

Chistoserdova, Ludmila, Marina G. Kalyuzhnaya, and Mary E. Lidstrom. 2009. “The Expanding World of Methylotrophic Metabolism.” *Annual review of microbiology* 63: 477–99.

Darling, Aaron E et al. 2014. “PhyloSift: Phylogenetic Analysis of Genomes and Metagenomes.” *PeerJ* 2: e243. https://peerj.com/articles/243.

Eiler, Alexander et al. 2015. “Tuning Fresh: Radiation through Rewiring of Central Metabolism in Streamlined Bacteria.” *The ISME Journal* (January 2016): 1–13. http://dx.doi.org/10.1038/ismej.2015.260.

Field, Dawn et al. 2006. “Open Software for Biologists: From Famine to Feast.” *Nature Biotechnology* 24(7): 801–3.

Ghylin, Trevor W. et al. 2014. “Comparative Single-Cell Genomics Reveals Potential Ecological Niches for the Freshwater acI Actinobacteria Lineage.” *The ISME journal* 8(12): 2503–16.

Gimmler, Anna et al. 2016. “The Tara Oceans Voyage Reveals Global Diversity and Distribution Patterns of Marine Planktonic Ciliates.” *Nature Publishing Group* (April): 1–13. http://dx.doi.org/10.1038/srep33555.

Giroldo, Danilo, Armando Augusto, and Henriques Vieira. 2005. “Polymeric and Free Sugars Released by Three Phytoplanktonic Species from a Freshwater Tropical Eutrophic Reservoir.” *Journal of Plankton Research* 27(7): 695–705.

Hall, Michael W et al. 2017. “Ananke : Temporal Clustering Reveals Ecological Dynamics of Microbial Communities.” *PeerJ* 5(e3812): 1–19.

Hamilton, Joshua J et al. 2017. “Metabolic Network Analysis and Metatranscriptomics Reveal Auxotrophies and Nutrient Sources of the Cosmopolitan Freshwater Microbial Lineage acI.” 2(4): 1–13.

Hanson, T E, and F R Tabita. 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.” *Proceedings of the National Academy of Sciences of the United States of America* 98(8): 4397–4402. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=31846&tool=pmcentrez&rendertype=abstract.

He, Shaomei et al. 2017. “Ecophysiology of Freshwater Verrucomicrobia Inferred from Metagenome-Assembled Genomes.” *mSphere* 2(5): 1–17.

Igarashi, Kazuei, and Keiko Kashiwagi. 1999. “Polyamine Transport in Bacteria and Yeast.” *Biochem. J.* 344: 633–42.

Jorgenson, Niels OG et al. 1998. “Effects of Sunlight on Occurrence and Bacterial Turnover of Specific Carbon and Nitrogen Compounds in Lake Water.” *FEMS Microbiology Ecology* 25: 217–27.

Juttner, F, and T Matuschek. 1977. “The Release of Low Molecular Weight Compounds by the Phytoplankton in an Eutrophic Lake.” *Water* 12: 251–55.

Kalyuzhnaya, Marina G et al. 2012. “Novel Methylotrophic Isolates from Lake Sediment, Description of Methylotenera Versatilis Sp. Nov. and Emended Description of the Genus Methylotenera.” *International journal of systematic and evolutionary microbiology* 62(Pt 1): 106–11. http://www.ncbi.nlm.nih.gov/pubmed/21335496.

Kanao, Tadayoshi et al. 2002. “Characterization of Isocitrate Dehydrogenase from the Green Sulfur Bacterium Chlorobium Limicola: A Carbon Dioxide-Fixing Enzyme in the Reductive Tricarboxylic Acid Cycle.” *European Journal of Biochemistry* 269(7): 1926–31.

Kang, Dongwan D, Jeff Froula, Rob Egan, and Zhong Wang. 2015. “MetaBAT, an Efficient Tool for Accurately Reconstructing Single Genomes from Complex Microbial Communities.” *PeerJ* 3: e1165. https://peerj.com/articles/1165.

Kara, Emily L et al. 2013. “A Decade of Seasonal Dynamics and Co-Occurrences within Freshwater Bacterioplankton Communities from Eutrophic Lake Mendota, WI, USA.” *The ISME journal* 7(3): 680–84. http://www.ncbi.nlm.nih.gov/pubmed/23051691 (February 11, 2014).

Karhunen, Jatta, Lauri Arvola, Sari Peura, and Marja Tiirola. 2013. “Green Sulphur Bacteria as a Component of the Photosynthetic Plankton Community in Small Dimictic Humic Lakes with an Anoxic Hypolimnion.” *Aquatic Microbial Ecology* 68: 267–72.

Latypova, Ekaterina et al. 2010. “Genetics of the Glutamate-Mediated Methylamine Utilization Pathway in the Facultative Methylotrophic Beta-Proteobacterium Methyloversatilis Universalis FAM5.” 75(December 2009): 426–39.

Li, Heng, and Richard Durbin. 2010. “Fast and Accurate Long-Read Alignment with Burrows-Wheeler Transform.” *Bioinformatics* 26(5): 589–95.

Linz, Alexandra M. et al. 2017. “Bacterial Community Composition and Dynamics Spanning Five Years in Freshwater Bog Lakes.” *mSphere* 2(3): 1–15.

Luo, Ruibang et al. 2012. “SOAPdenovo2: An Empirically Improved Memory-Efficient Short-Read de Novo Assembler.” *GigaScience* 1(1): 18. http://www.gigasciencejournal.com/content/1/1/18%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3626529&tool=pmcentrez&rendertype=abstract.

Markowitz, Victor M. et al. 2012. “IMG: The Integrated Microbial Genomes Database and Comparative Analysis System.” *Nucleic Acids Research* 40(D1): 115–22.

Martinez-Garcia, Manuel et al. 2012. “High-Throughput Single-Cell Sequencing Identifies Photoheterotrophs and Chemoautotrophs in Freshwater Bacterioplankton.” *The ISME Journal* 6(1): 113–23. http://dx.doi.org/10.1038/ismej.2011.84.

McGowan, Suzanne et al. 2016. “Long-Term Perspectives on Terrestrial and Aquatic Carbon Cycling from Palaeolimnology.” *WIREs Water* 3: 211–34.

Miles, R J. 1992. “Review Article Catabolism in *Mollicutes*.” *Journal of General Microbiology* 138(1 992): 1773–83.

Mou, Xiaozhen et al. 2011. “Metatranscriptomic Signature of Exogenous Polyamine Utilization by Coastal Bacterioplankton.” 3: 798–806.

Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. “APE: Analyses of Phylogenetics and Evolution in R Language.” *Bioinformatics* 20(2): 289–90.

Parks, Donovan H et al. 2015. “CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes.” *Genome Research* 25(7).

Paver, Sara F et al. 2017. “Temporal Patterns in Glycolate-Utilizing Bacterial Community Composition Correlate with Phytoplankton Population Dynamics in Humic Lakes Linked References Are Available on JSTOR for This Article : Temporal Patterns in Glycolate-Utilizing Bacterial Communi.” 60(2): 406–18.

Peters, John W. et al. 2015. “[FeFe]- and [NiFe]-Hydrogenase Diversity, Mechanism, and Maturation.” *Biochimica et Biophysica Acta - Molecular Cell Research* 1853(6): 1350–69. http://dx.doi.org/10.1016/j.bbamcr.2014.11.021.

Peura, Sari et al. 2012. “Distinct and Diverse Anaerobic Bacterial Communities in Boreal Lakes Dominated by Candidate Division OD1.” *ISME journal* 6: 1640–52.

Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. 2010. “FastTree 2 - Approximately Maximum-Likelihood Trees for Large Alignments.” *PLoS ONE* 5(3).

Remsen, Charles C, Edward J Carpenter, and Brian W Schroeder. 1972. “Competition for Urea among Estuarine Microorganisms.” *Ecological Society of America* 53(5): 921–26.

Rinke, Christian et al. 2013. “Insights into the Phylogeny and Coding Potential of Microbial Dark Matter.” *Nature* 499(7459): 431–37. http://www.ncbi.nlm.nih.gov/pubmed/23851394.

Salcher, Michaela M, Stefan M Neuenschwander, Thomas Posch, and Jakob Pernthaler. 2015a. “The Ecology of Pelagic Freshwater Methylotrophs Assessed by a High-Resolution Monitoring and Isolation Campaign.” 9(11): 2442–53. http://dx.doi.org/10.1038/ismej.2015.55.

———. 2015b. “The Ecology of Pelagic Freshwater Methylotrophs Assessed by a High-Resolution Monitoring and Isolation Campaign.” *The ISME Journal*: 1–12. http://www.nature.com/doifinder/10.1038/ismej.2015.55.

Sommer, Daniel D, Arthur L Delcher, Steven L Salzberg, and Mihai Pop. 2007. “Minimus: A Fast, Lightweight Genome Assembler.” *BMC bioinformatics* 8: 64.

Tang, Kuo Hsiang, and Robert E. Blankenship. 2010. “Both Forward and Reverse TCA Cycles Operate in Green Sulfur Bacteria.” *Journal of Biological Chemistry* 285(46): 35848–54.

Team, R Core. 2015. “R: A Language and Environment for Statistical Computing.” *R Foundation for Statistical Computing, Vienna, Austria*: http://www.R-project.org/.

Varghese, Neha J et al. 2015. “Microbial Species Delineation Using Whole Genome Sequences.” *Nucleic acids research* 43(14): gkv657-. http://nar.oxfordjournals.org/content/early/2015/07/06/nar.gkv657.full.

Wickham, Hadley. 2007. 21(12) Journal of Statistical Software *Reshaping Data with the Reshape Package*.

———. 2009. “ggplot2: Elegant Graphics for Data Analysis.” *Spring-Verlag New York*.

Williamson, Craig E., Walter Dodds, Timothy K. Kratz, and Margaret A. Palmer. 2008. “Lakes and Streams as Sentinels of Environmental Change in Terrestrial and Atmospheric Processes.” *Frontiers in Ecology and the Environment* 6(5): 247–54.

Yin, Yanbin et al. 2012. “DbCAN: A Web Resource for Automated Carbohydrate-Active Enzyme Annotation.” *Nucleic Acids Research* 40(W1): 445–51.

Zheng, Yanning et al. 2018. “A Pathway for Biological Methane Production Using Bacterial Iron-Only Nitrogenase.” *Nature Microbiology* 4. http://www.nature.com/articles/s41564-017-0091-5.

## Supplemental

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lake Mendota | Trout Bog Epilimnion | Trout Bog Hypolimnion |
| Chloride (ppm) | 39.87 (7.44) | 0.22 (0.07) | 0.28 (0.08) |
| Calcium (ppm) | 31.90 (5.04) | 1.38 (0.24) | 1.84 (0.35) |
| Magnesium (ppm) | 32.93 (2.72) | 0.39 (0.06) | 0.45 (0.06) |
| Sodium (ppm) | 19.48 (1.69) | 0.22 (0.07) | 0.25 (0.06) |
| Potassium (ppm) | 3.27 (0.28) | 0.64 (0.17) | 0.75 (0.17) |
| Iron (ppm) | 0.00 (0.01) | 0.31 (0.10) | 0.47 (0.09) |
| Manganese (ppm) | 0.00 (0.01) | 0.08 (0.07) | 0.09 (0.11) |

**Table S1. Additional chemical concentrations in our study sites.**