# Linkages between freshwater carbon nutrient cycles revealed through time series metagenomics

Alexandra M. Linz1\*, Shaomei He1, Sarah L. R. Stevens1, Karthik Anantharaman1, Joshua J. Hamilton2, Robin R. Rohwer3, Rex R. Malmstrom4, Stefan Bertilsson5, Katherine D. McMahon1,3

1Department of Bacteriology, University of Wisconsin – Madison, 2Department of Biochemistry, University of Wisconsin – Madison, 3Department of Civil and Environmental Engineering, University of Wisconsin – Madison, 4Department of Energy Joint Genome Institute, 5Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University

\*Corresponding author: Alexandra Linz, Department of Bacteriology, University of Wisconsin 1550 Linden Drive Room 5525 Madison, WI 53706, Email: [amlinz@wisc.edu](mailto:amlinz@wisc.edu)

Abstract word count: 275

Article word count: 5,028 (including citations)

Running title: Linkages between freshwater nutrient cycles

# Abstract

Because microbes are responsible for much of the carbon and nutrient cycling in freshwater, metabolic processes at the microbial scale influence ecosystem-level functions. One approach to predicting the metabolic capabilities of microbial communities is searching for functional marker genes in metagenomes. However, this method does not provide information about the populations carrying these marker genes, such as detailed taxonomy and co-occurrence with other metabolic traits at an organism-scale. In this study, we combine a functional marker gene analysis with pathway prediction of metagenome-assembled genomes from a metagenomic time series to ask how nutrient cycles are linked in freshwater bacterioplankton. Comparing our results across Lake Mendota, a eutrophic lake, and Trout Bog, a humic lake, revealed both similarities and differences in microbial nutrient cycling. Phototrophy, carbon fixation, and nitrogen fixation pathways co-occurred in Cyanobacteria populations in Lake Mendota and in Chlorobiales in Trout Bog. Using our time series, we show that the abundance of Cyanobacteria MAGs strongly correlated to functional marker genes for nitrogen fixation in several years. Genes encoding steps in the nitrogen and sulfur cycles varied in abundance and taxonomy by lake, presumably reflecting the availability and composition of inorganic nutrients in these systems. We were also able to identify which populations contained the greatest density and diversity of genes encoding glycoside hydrolases. Populations with many glycoside hydrolases also encoded pathways for sugar degradation. Using both genomes and marker genes, we are better able to link function to specific taxonomic groups in our metagenomic time series. These genome-enabled insights and others presented here enable a more detailed understanding of freshwater microbial nutrient cycling.

# Introduction

Lakes collect nutrients from surrounding terrestrial ecosystems (Williamson et al., 2008), placing lakes as “hotspots” for carbon and nutrient cycling in the landscape (Butman et al., 2015). Much of this biogeochemical cycling is performed by freshwater microbes. We have learned much about freshwater microbes through previous research has revealed high levels of diversity and change over time in freshwater microbial communities (Allgaier & Grossart, 2006), the geographic distribution of freshwater taxa (Šimek et al., 2010), the distribution of functional marker genes (Peura et al., 2012, 2015; Ramachandran & Walsh, 2015; Eiler et al., 2016), and substrate use capabilities in specific phylogenetic groups (Salcher, Posch & Pernthaler, 2013). However, organism-level information about microbial metabolisms is currently not incorporated well into conceptual models of freshwater nutrient cycling.

Although aquatic microbes are often classified exclusively as decomposers or phytoplankton, their role and relative importance in the food chain is now recognized as distinct and complex (Pomeroy & Wiebe, 1988). Dissolved organic carbon (DOC) is produced at every trophic level, but this carbon is often not in a form directly available to be consumed by secondary or tertiary trophic levels. Instead, microbes are responsible for processing this DOC, producing biomass, and subsequently being consumed. This process of maintaining lost DOC within the food web is known as the “microbial loop” (Azam et al., 1983), although aquatic microbes respire much of the DOC to CO2. In some systems, microbial respiration is thought to exceed primary production, resulting in the release of excess of CO2 to the atmosphere (del Giorgio, Cole & Cimbleris, 1997). Inorganic compounds can be used as nitrogen and sulfur sources or provide energy to chemolithotrophs that is used by other trophic levels. Microbial conversions of inorganic compounds are often just as crucial to freshwater nutrient cycling as the degradation of organic compounds.

Previously, we used time series metagenomics to assemble nearly 200 medium to high quality (metagenome-assembled genomes) MAGs from two temperate lakes: Lake Mendota, a highly productive eutrophic lake, and Trout Bog, a humic bog lake (Bowers et al., 2017). This dataset has been used to study genome-wide diversity 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>), their previous 16S rRNA gene amplicon time series analyses (Hall et al., 2017; Linz et al., 2017), and their contrasting limnological attributes (Table 1, Table S1).

While gene-centric methods are one method that can be used to infer the potential metabolic processes occurring at a community level, these approaches cannot address the critical role of 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 metabolism 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.

# Methods

Samples were collected from Lake Mendota and Trout Bog as previously described (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 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-µm 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 previously described (Bendall et al., 2016; Roux et al., 2017), 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) (Data S1). 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 merge read lengths of 150-290 bp (Magooc & Salzberg, 2011). 16S rRNA amplicon iTag sequencing was also performed on these samples. Samples from Trout Bog were sequenced on a 454 GS FLX-Titanium; the V6-V8 region was targeted using primers 926F and 1392R (Engelbrektson et al., 2010). Samples from Lake Mendota were sequenced on an Illumina MiSeq, and the V4 region was targeted using primers 525F and 806F (Caporaso et al., 2012). 16S rRNA amplicon data was analyzed in mothur and classified using the freshwater-specific TaxAss workflow as previously described (Schloss et al., 2009; Hall et al., 2017; Linz et al., 2017).

To analyze functional marker genes in the unassembled, unpooled metagenomes, we used a custom database of reference protein sequences (Data S2) (Karthik - any citation needed?) and identified open reading frames (ORFs) in our unassembled metagenomic time series using Prodigal (Hyatt et al., 2010). This analysis was conducted on merged reads. The protein sequences and ORFs were compared using BLASTx (Camacho et al., 2009), with a cutoff of 30% ID across the length of the ORF. Significant differences in gene frequency between sites were tested using LEfSE (Segata et al., 2012). 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.

To recover MAGs, metagenomic reads were pooled by lake and layer and then assembled as previously described (Bendall et al., 2016; Roux et al., 2017). In Trout Bog, this assembly was performed using SOAPdenovo at various k-mer sizes (Luo et al., 2012) and the resulting contigs were combined using Minimus (Sommer et al., 2007). In Lake Mendota, merged reads were assembled using Ray v2.20 with a single k-mer size (Boisvert et al., 2012). Contigs from the combined assemblies were binned using MetaBAT (Kang et al., 2015) and reads from unpooled metagenomes were mapped to the assembled contigs using the Burrows-Wheeler Aligner (Li & Durbin, 2010), allowing time-series resolved binning (Table S2). DOE JGI’s Integrated Microbial Genome (IMG) database tool (https://img.jgi.doe.gov/mer/) (Markowitz et al., 2012) 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 (Rinke et al., 2013; Parks et al., 2015), and MAGs were classified using Phylosift (Darling et al., 2014).

Only MAGs at least 50% complete with less than 10% estimated contamination (meeting the MIMARKS definition of at least a medium quality MAG) were included in this study (Bowers et al., 2017). 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 describe (He et al., 2017). To score presence, 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. 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) (Yin et al., 2012). 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 first tested whether functional marker genes identified in the unassembled merged metagenomic reads appeared more frequently in one lake or layer compared to the others. 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 the comparison illogical. Many genes differed significantly by site, indicating contrasting gene content between lakes and layers (Data S3). To further infer differences in microbial metabolism from these data, 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.

This analysis revealed that RuBisCo, the marker gene for carbon fixation via the Calvin-Benson-Bassham (CBB) cycle, was most frequently observed in the epilimnion of Trout Bog (Figure 1, Table S3). Citrate lyase, the marker for carbon fixation via the reverse TCA cycle, was observed more often in the hypolimnion of Trout Bog than in the two epilimnion sites. Most nitrogen cycling genes differed significantly across all three sites, with the Trout Bog hypolimnion typically having greater numbers of hits for these genes. Sulfur cycling genes showed fewer significant differences between sites, with sulfur oxidation (SOX) and sulfide reduction (sulfide quinone reductase) genes more abundant in Trout Bog compared to Mendota. These contrasts in the abundance of functional marker genes suggested significant differences in the metabolisms of microbial communities between lakes and layers.

## Overview of the MAGs

While our functional marker gene analysis demonstrated significant differences in gene content between our study sites, it cannot always provide reliable information about the identity of the microorganisms carrying these genes or about the co-occurrence of these genes within the same populations. Different evolutionary rates and horizontal gene transfer make inferring taxonomy from a single gene problematic. To address these questions, we assembled MAGs from the same metagenomic time series and predicted metabolic pathways present in freshwater populations based on genomic content.

A total of 193 medium to high quality bacterial metagenome assembled genomes (MAGs) were recovered from the three combined time series metagenomes in Trout Bog and Lake Mendota (Data S5). These genomes ranged in estimated completeness from 50 to 99% based on CheckM estimates (Parks et al., 2015). Of the 193 MAGs, 99 were recovered from Lake Mendota, 31 were recovered from the Trout Bog epilimnion, and 63 were recovered from the Trout Bog hypolimnion. Several MAGs from the Trout Bog epilimnion and hypolimnion appeared to be from the same population based on average nucleotide identities greater than 99% using JGI’s ANI calculator (Data S6) (Varghese et al., 2015). 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 (Price, Dehal & Arkin, 2010) using whole genome alignments (Figure S1). The tree is not intended to infer detailed evolutionary history, but to provide an overall picture of similarity between genomes.

The phylum-level assignments of our MAGs were largely consistent with the classifications of 16S rRNA gene amplicon sequencing results averaged across the time series, presumably because MAGs were more likely to be recovered from abundant populations in the community (Figure 2, Data S4). Some taxa such as Tenericutes, Ignavibacteria, Epsilonproteobacteria, and Chlamydiae were represented by MAGs, but not by 16S rRNA sequences. Chlorobi appeared overrepresented by MAG coverage compared to 16S rRNA gene counts, while Proteobacterial groups appeared to comprise more of the community by 16S than by MAG coverage. These discrepancies could be explained by bias in the 16S primer sets (Hong et al., 2009) difference in *rRNA* copy number, or assembly bias in MAG recovery. The observed community compositions are consistent with other 16S-based studies for these sites (Hall et al., 2017; Linz et al., 2017). Detecting similar phyla using both methods suggests that our MAG datasets are largely representative of the resident microbial communities.

## Nitrogen Cycling

Nitrogen availability is an important factor structuring freshwater microbial communities. Our analysis of nitrogen-related marker genes revealed many significant differences in the abundance of these markers in Lake Mendota, the Trout Bog epilimnion, and the Trout Bog hypolimnion (Figure 1). Genes encoding nitrogenase subunits were observed most frequently in metagenomes from the hypolimnion of Trout Bog, followed by the epilimnion of Trout Bog, and lastly Lake Mendota. As the nitrogenase enzyme is inhibited by oxygen, nitrogen fixation may be most favorable in the anoxic layer of Trout Bog.

Analysis of genes encoding nitrogenases in the MAGs showed differences in the taxonomy of putative diazotrophs between the two ecosystems (Figure 3, Figure S1). In Lake Mendota, every MAG classified as Cyanobacteria contained genes encoding the nitrogen fixation pathway, with a handful of additional putative diazotrophs assigned to Betaproteobacteria, and Gammaproteobacteria. This strong association between Cyanobacteria and nitrogen fixation genes supports the documented links between cyanobacterial bloom toxicity and nitrogen fixation in Lake Mendota (Beversdorf, Miller & McMahon, 2013). MAGs containing genes encoding nitrogen fixation were more phylogenetically diverse in Trout Bog, including 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.

Marker genes for denitrification pathways show a similar distribution as genes encoding nitrogenase subunits; they were observed most frequently in metagenomes from the Trout Bog hypolimnion, except for nitrous oxide reductase, which was most frequently found in Lake Mendota. Putative denitrification pathways were widespread in MAGs from both lakes (Figure 3). Genes encoding ureases were not found significantly more often in any site, and urea degradation pathways were predicted in MAGs from both lakes. This is consistent with research showing that urea is a common nitrogen source for bacteria in multiple freshwater types (Remsen, Carpenter & Schroeder, 1972; Jorgenson et al., 1998; Berman & Bronk, 2003).

Genes encoding the biosynthesis and degradation of polyamines such as spermidine and putrescine were prevalent in many diverse MAGs from both lakes, including Actinobacteria as has been previously observed (Ghylin et al., 2014). While there is some evidence for the importance of polyamines in aquatic systems (Mou et al., 2011), the ecological role of these compounds in freshwater is not yet fully resolved. Polyamines are known to play a critical but poorly understood role in bacterial metabolism (Igarashi & Kashiwagi, 1999), and the exchange of these nitrogen compounds between populations may be a factor structuring freshwater microbial communities. One additional source of polyamines is higher trophic levels such as fish or zooplankton, as these compounds can result from bacterial decomposition of amino acids (Al Bulushi et al., 2009). 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 ecosystems.

Properties of genomes themselves may also provide information about nitrogen limitation (Data S5) (Bragg, 2011). We observed a bias in MAGs from Trout Bog towards encoding amino acids with 1% less nitrogen compared to MAGs from Lake Mendota (Data S5). Although this difference is small, it was significant using a Wilcoxon rank sum test (p = 0.02). This amino acid bias suggests that conditions in Trout Bog may lead to stronger selection for organisms encoding nitrogen poor proteins compared to the conditions 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 from the surrounding agriculture 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 sphagnum moss growing its edges.

## Sulfur Cycling

Sulfur is another essential element in freshwater that is processed by microbes. Our marker gene analysis demonstrated that genes encoding sulfide:quinone reductase and SOX sulfur oxidation were significantly more abundant in Trout Bog compared to Lake Mendota, with no significant differences between the layers of Trout Bog (Figure 1). Genes encoding sulfite reductase were the least abundant sulfur cycling marker gene 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 (Holkenbrink et al., 2011), 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 3). The relative ratios of assimilatory to dissimilatory sulfate reduction predicted pathways suggest 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 (Bowles et al., 2014), although sulfate reduction could also be occurring in Lake Mendota’s hypolimnion.. Sulfur oxidation pathways were observed in MAGs classified as Betaproteobacteria from both of our lakes and Epsilonproteobacteria in the Trout Bog Hypolimnion.

## Phototrophy

Primary production is a critical component of the freshwater carbon cycle. Therefore, we looked at potential routes of primary production within the MAGs, expecting to find differences between our two study sites based on the observed contrasts in the functional marker gene analysis. 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, MAGs encoding phototrophy 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 & Blankenship, 2010). The marker gene for this pathway (citrate lyase) was significantly more abundant in the Trout Bog hypolimnion than in the epilimnion, consistent with its presence in a strictly anaerobic lineage (Figure 1). Both photoautotrophs also contained genes potentially encoding nitrogen fixation. The linkages in these populations are especially interesting given that both photoautotrophs are relatively abundant community members in their respective lakes.

Genes annotated as the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) large subunit (*rbcL*), a 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 cultured representatives of Chlorobiales, and a RuBisCo-like gene found in isolates of *Chlorobium* was associated with sulfur metabolism and oxidative stress (Hanson & Tabita, 2001). 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 gene encodes a function other than carbon fixation in the Chlorobiales MAGs. Instead, the high abundances of the RuBisCO marker gene identified in the Trout Bog epilimnion (Figure 1) are more likely derived from the eukaryotic algae and Cyanobacteria detected in our 16S rRNA gene survey than from the strictly anaerobic Chlorobiales.

The potential for photoheterotrophy via the aerobic anoxygenic phototrophic pathway was identified in several MAGs from both lakes, especially from epilimnia, based on the presence of genes annotated as *pufABCLMX, puhA,* and *pucAB* encoding the core reaction center RC-LH1(Martinez-Garcia et al., 2012). Betaproteobacteria and Gammaproteobacteria, particularly MAGs classified as Burkholderiales, most often contained these genes, although they were not broadly shared across the phylum (Fig. 3). Aerobic anoxygenic phototrophy has previously been associated with these taxa (Martinez-Garcia et al., 2012). An Acidobacterial 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 (Martinez-Garcia et al., 2012). We observed genes encoding rhodopsins in MAGs from both lakes, but more frequently in Actinobacteria and Bacteroidetes MAGs from Lake Mendota. Trout Bog, especially the hypolimnion, harbored much less diversity and a lower frequency of 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). While organic carbon in freshwater is often classified as either autochthonous and allochthonous carbon, this distinction is less relevant for organotrophic bacteria. For example, there is substantial overlap in the molecular composition of algal exudates, cellulose degradation intermediates, and photochemical degradation products (Bertilsson & Tranvik, 1998; Ramanan et al., 2015). One-carbon compounds such as methane are produced in the lake (therefore autochthonous), but they are often 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.

While degradation of high-complexity carbon compounds may require specialized enzymes, their wide availability and high yield of sugars makes this 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), enzymes that break 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 available carbon sources (He et al., 2017). Here, we expand this analysis of glycoside hydrolases to the all MAGs in our dataset to identify differences in how populations from our two study sites degrade complex carbohydrates.

Calculating the coding density of GHs– the percentage of coding regions in a MAG annotated as a GH – immediately revealed differences between the Trout Bog and the Lake Mendota MAGs, and even between MAGs from the epilimnion and hypolimnion of Trout Bog (Fig. 4). The MAGs with the highest GH 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 GHs in Lake Mendota and the Trout Bog epilimnion, but the others did not. There were several orders with GHs 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, these MAGs from both lakes also contain putative degradation pathways for a variety of sugars (Figure 3). The GH diversity, an indicator of the number of substrates an organism can utilize, was significantly correlated with GH coding density at the order level (r2= 0.39, p = 4.5x10-8).

Genes encoding for several GH families were identified in MAGs from Lake Mendota and both layers of Trout Bog. Starting with the most abundant, 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 (He et al., 2017). Lake Mendota contained unique GHs belonging to the family GH13 (alpha-glucoside). The only unique GH in the Trout Bog epilimnion was GH62, a putative arabinofuranosidase. The hypolimnion contained many more unique enzymes than Lake Mendota or the epilimnion of Trout Bog, 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 in Trout Bog’s hypolimnion suggested differences between their GH profiles, presumably correlated to the diversity and complexity of their carbon sources.

## 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 to reduce carbon demand, were observed in Alphaproteobacteria and Chlamydiae in Lake Mendota and Acidobacteria and Betaproteobacteria in Trout Bog.

As oxidative phosphorylation is an important part of central metabolism for aerobic bacteria, we investigated the types of cytochrome oxidases encoded in our MAGs (Figure 3). 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 (Gong et al., 2018), 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 the hypolimnion of Trout Bog, while cytochrome aa3-600 was found only in MAGs classified as Bacteroidetes and Betaproteobacteria in the Trout Bog epilimnion. Cytochrome o oxidase was observed only in a Chlamydia MAG from Lake Mendota. Alternative complex III was identified in MAGs of Verrucomicrobia in both lakes, in Acidobacteria in Trout Bog (both layers), and in Bacteroidetes and Planctomycetes in Lake Mendota.

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 MAGs from Trout Bog’s hypolimnion (Figure 3, Table S3), consistent with their previously identified presence in anaerobic, often fermentative bacteria (Peters et al., 2015) and significantly higher observations of marker genes encoding 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 the Mendota and Trout Bog epilimnion and group 2b in the Trout Bog epilimnion and hypolimnion. Genes encoding these hydrogenases were widespread in MAGs from the hypolimnion of Trout Bog, found only in Chlorobiales MAGs in the epilimnion of Trout Bog, and more 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 (Giroldo, Augusto & Vieira, 2005; Ramanan et al., 2015). 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 in Lake Mendota, and members of Bacteroidetes, Ignavibacteria, and Verrucomicrobia in 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 (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 & Kent, 2017), were identified in Cyanobacteria and Betaproteobacteria in Lake Mendota and in Acidobacteria, Verrucomicrobia, Alpha-, Beta-, Gamma-, and Epsilonproteobacteria in 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 and Methylotenera (now merged with Nitrosomonadales (Boden, Hutt & Rae, 2017)), while MAGs from Methylococcales were potential methane degraders based on the presence of genes encoding methane monooxygenase. MAGs of Methylococcales in Trout Bog also encoded the pathway for nitrogen fixation, consistent with reports of nitrogen fixation in cultured isolates of this taxon (Bowman, Sly & Stackebrandt, 1995). 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). Methylotrophy in cultured freshwater isolates from these taxa is well-documented (Kalyuzhnaya et al., 2011; Salcher et al., 2015); however, genes encoding methanol degradation were also identified in MAGs from taxa not typically associated with methylotrophy. These included MAGs classified as Burkholderiales, Rhizobiales, and Nitrosomonadales in Trout Bog. Given the rapid rate at which we are discovering methylotrophy in microorganisms not thought to be capable of this process, this finding is intriguing, but not surprising (Chistoserdova, Kalyuzhnaya & Lidstrom, 2009).

## Using MAGs to track population abundances over time

As our metagenomes comprise a time series, 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 5, A-E). We found that one Cyanobacterial 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 dominant Cyanobacterial population in each year (Figure 5, 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 Cyanobacterial MAG abundancemore 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 Cyanobacterial 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 in our analysis. These two years were also unusual in our time series - in 2008, extreme flooding events led to large Cyanobacterial blooms (Beversdorf et al., 2015) and in 2009, the invasive spiny water flea was first detected in Lake Mendota (Walsh, Carpenter & Vander Zanden, 2016). Still, our time series analysis demonstrates the utility of our datasets in linking metabolic function to specific taxonomic groups.

## Conclusions

Our analysis of functional marker genes indicated significant differences in microbial nutrient cycling between the Lake Mendota epilimnion, Trout Bog epilimnion, and Trout Bog hypolimnion. By combining these results with metabolic pathway prediction in MAGs, we were able to identify taxa encoding these metabolisms and co-occurrence of pathways within populations. We found that phototrophy, carbon fixation, and nitrogen fixation are linked in both lakes through the abundant phototrophs Cyanobacteria in Lake Mendota and Chlorobiales in Trout Bog. In Lake Mendota, nitrogen fixation was associated with Cyanobacteria, while MAGs encoding nitrogen fixation were more diverse in Trout Bog. In the sulfur cycle, assimilatory pathways were observed more frequently in the MAGs than dissimilatory pathways, suggesting a bias towards using sulfur compounds in biosynthesis rather than for energy mobilization. Analysis of genes annotated as glycoside hydrolases in the MAGs revealed the greatest density and diversity of these enzymes in the Trout Bog hypolimnion, potentially indicating a greater reliance on a diversity of complex carbon sources. This combined data from functional marker gene analysis and MAG pathway prediction provides insight into the complex metabolisms underpinning freshwater communities and how microbial processes scale to ecosystem functions.

### Acknowledgments

We thank the North Temperate Lakes and Lake Mendota Microbial Observatory field crews, UW-Trout Lake Station, the UW Center for Limnology, and the Global Lakes Ecological Observatory Network for field and logistical support. We acknowledge efforts by many McMahon laboratory undergraduate students and technicians whose work has been related to sample collection and DNA extraction. We thank Emily Stanley for insightful comments on an early draft of this manuscript. Finally, we personally thank the individual program directors and leadership at the National Science Foundation for their commitment to continued support of long-term ecological research.

We thank the Joint Genome Institute for supporting this work through the Community Sequencing Program (CSP 394), performing the bioinformatics, and providing technical support. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. K.D.M. acknowledges funding from the United States National Science Foundation Microbial Observatories program (MCB-0702395), the Long Term Ecological Research Program (NTL-LTER DEB- 1440297), and an INSPIRE award (DEB-1344254). A.M.L. was supported by a pre-doctoral fellowship provided by the University of Wisconsin – Madison Department of Bacteriology and by the National Science Foundation Graduate Research Fellowship Program under grant no. DGE-1256259 during this research.

# Figure and Table Legends

**Table 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.

**Figure 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 blue 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. LEfSe results for each gene are available in Data S3, and p-values of markers described in Figure 1 and elsewhere in the text are reported in Table S3.

**Figure 2. 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.

**Figure 3. Metabolisms in Lake Mendota vs. 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.

**Figure 4. 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 (r2 = 0.38, p = 4.5x10-8). 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.

**Figure 5. 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.

**Supplemental Legends**

**Table S1. Additional chemical measurements in our study sites.** Additional chemistry data were collected by NTL-LTER (<http://lter.limnology.wisc.edu>) 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.

**Data S1. Metagenome metadata**. This dataset includes information about the metagenomes used in this study including date collected, size in reads and base pairs, and their IMG Genome IDs (IMG Taxon ID).

**Data S2. Functional marker genes.** This dataset lists the TIGRFAM, COG, or PFAM IDs of sequences used as functional marker genes in this study.

**Table S2. 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 medium to high quality metagenome-assembled genomes.

**Data S3. Results of LEfSe analysis on functional marker genes**. The program LEfSe was used to detect significant differences in gene content between our study sites. The distinguish feature of LEfSe, the LDA effect score, is listed for each marker gene in this dataset.

**Table S3. 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.

**Data S4. 16S rRNA amplicon sequencing of our samples.** 16S sequencing was performed over the time series to assess community composition in our study sites. The resulting OTU tables and taxonomic classifications are presented here.

**Data S5. MAG metadata.** Information about the completeness, size, and taxonomy of our MAGs, as well as their IMG OIDs, are presented here. Amino acid use was calculated based on the average number of nitrogen atoms translated gene sequences.

**Data S6. Average nucleotide identity between MAGs.** Average nucleotide identity (ANI) was calculated between all MAGs in our dataset. MAGs with extremely high ANIs (>97%) are likely from the same populations. An ANI value of “0” indicates that no portions of the genomes aligned.

**Figure S1. 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.

# References

Allgaier M., Grossart H-P. 2006. Diversity and Seasonal Dynamics of Actinobacteria Populations in Four Lakes in Northeastern Germany. *Applied and Environmental Microbiology* 72:3489–3497. DOI: 10.1128/AEM.72.5.3489.

Azam F., Fenchel T., Field JG., Gray JC., Meyer-Reil LA., Thingstad F. 1983. The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series* 10:257–264. DOI: 10.3354/meps010257.

Bendall ML., Stevens SLR., Chan L., Malfatti S., Schwientek P., Tremblay J., Schackwitz W., Martin J., Pati A., Bushnell B., Froula J., Kang D., Tringe SG., Bertilsson S., Moran MA., Shade A., Newton RJ., McMahon KD., Malmstrom RR. 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *The ISME Journal* 10:1589–1601. DOI: 10.1038/ismej.2015.241.

Berman T., Bronk DA. 2003. Dissolved organic nitrogen: A dynamic participant in aquatic ecosystems. *Aquatic Microbial Ecology* 31:279–305. DOI: 10.3354/ame031279.

Bertilsson S., Tranvik LJ. 1998. Photochemically produced carboxylic acids as substrates for freshwater bacterioplankton. *Limnology and Oceanography* 43:885–895. DOI: 10.4319/lo.1998.43.5.0885.

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* 10:1–18. DOI: 10.1371/journal.pone.0125353.

Beversdorf LJ., Miller TR., McMahon KD. 2013. The Role of Nitrogen Fixation in Cyanobacterial Bloom Toxicity in a Temperate , Eutrophic Lake. *PLOS ONE* 8:1–11. DOI: 10.1371/journal.pone.0056103.

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. *International Journal of Systematic and Evolutionary Microbiology* 67:1191–1205. DOI: 10.1099/ijsem.0.001927.

Boisvert S., Raymond F., Godzaridis É., Laviolette F., Corbeil J. 2012. Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biology* 13:1–13. DOI: 10.1186/gb-2012-13-12-r122.

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 WT., 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., 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. *Nature Biotechnology* 35:725–731. DOI: 10.1038/nbt.3893.

Bowles MW., Mogollon JM., Kasten S., Zabel M., Hinrichs K-U. 2014. Global rates of marine sulfate reduction and implications for sub-sea-floor metabolic activities. *Science Express Reports*. DOI: 10.1038/35351.

Bowman JP., Sly LI., Stackebrandt E. 1995. The phylogenetic position of the family Methylococcaceae. *International Journal of Systematic Bacteriology* 45:182–5. DOI: 10.1099/00207713-45-3-622a.

Bragg JG. 2011. How Prochlorococcus bacteria use nitrogen sparingly in their proteins. *Molecular Ecology* 20:27–28. DOI: 10.1111/j.1365-294X.2010.04915.x.

Al Bulushi I., Poole S., Deeth HC., Dykes GA. 2009. Biogenic amines in fish: Roles in intoxication, spoilage, and nitrosamine formation-A review. *Critical Reviews in Food Science and Nutrition* 49:369–377. DOI: 10.1080/10408390802067514.

Butman D., Stackpoole S., Stets E., McDonald CP., Clow DW., Striegl RG. 2015. Aquatic carbon cycling in the conterminous United States and implications for terrestrial carbon accounting. *Proceedings of the National Academy of Sciences*:1–6. DOI: 10.1073/pnas.1512651112.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden TL. 2009. BLAST plus : architecture and applications. *BMC Bioinformatics* 10:1–9. DOI: Artn 421\nDoi 10.1186/1471-2105-10-421.

Caporaso JG., Lauber CL., Walters WA., Berg-Lyons D., Huntley J., Fierer N., Owens SM., Betley J., Fraser L., Bauer M., Gormley N., Gilbert JA., Smith G., Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal* 6:1621–1624. DOI: 10.1038/ismej.2012.8.

Chistoserdova L., Kalyuzhnaya MG., Lidstrom ME. 2009. The Expanding World of Methylotrophic Metabolism. *Annual Review of Microbiology* 63:477–499. DOI: 10.1146/annurev.micro.091208.073600.The.

Darling AE., Jospin G., Lowe E., Matsen FA., Bik HM., Eisen JA. 2014. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2:e243. DOI: 10.7717/peerj.243.

Eiler A., Mondav R., Sinclair L., Fernandez-Vidal L., Scofield D., Scwientek P., Martinez-Garcia M., Torrents D., McMahon KD., Andersson SGE., Stepanauskas R., Woyke T., Bertilsson S. 2016. Tuning fresh: radiation through rewiring of central metabolism in streamlined bacteria. *The ISME Journal* 10:1–13. DOI: 10.13140/RG.2.1.1968.9040.

Engelbrektson AL., 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. *The ISME Journal* 4:642.

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. *The ISME Journal* 8:2503–16. DOI: 10.1038/ismej.2014.135.

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

Giroldo D., Augusto A., Vieira H. 2005. Polymeric and free sugars released by three phytoplanktonic species from a freshwater tropical eutrophic reservoir. *Journal of Plankton Research* 27:695–705. DOI: 10.1093/plankt/fbi043.

Gong X., Garcia-Robledo E., Revsbech N-P., Schramm A. 2018. Gene Expression of Terminal Oxidases in Two Marine Bacterial Strains Exposed to Nanomolar Oxygen Concentrations. *FEMS Microbiology Ecology*. DOI: 10.1093/femsec/fiy072/4983120.

Hall MW., Rohwer RR., Perrie J., Mcmahon KD., Beiko RG. 2017. Ananke : temporal clustering reveals ecological dynamics of microbial communities. *PeerJ* 5:1–19. DOI: 10.7717/peerj.3812.

Hamilton JJ., Garcia SL., Brown BS., Oyserman BO., Moya-Flores F., Bertilsson S., McMahon Katherine D. 2017. Metabolic Network Analysis and Metatranscriptomics Reveal Auxotrophies and Nutrient Sources of the Cosmopolitan Freshwater Microbial Lineage acI. *mSystems* 2:1–13.

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. *Proceedings of the National Academy of Sciences* 98:4397–4402. DOI: 10.1073/pnas.081610398.

He S., Stevens SL., 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* 2:1–17.

Holkenbrink C., Barbas SO., Mellerup A., Otaki H., Frigaard NU. 2011. Sulfur globule oxidation in green sulfur bacteria is dependent on the dissimilatory sulfite reductase system. *Microbiology* 157:1229–1239. DOI: 10.1099/mic.0.044669-0.

Hong S., Bunge J., Leslin C., Jeon S., Epstein SS. 2009. Polymerase chain reaction primers miss half of rRNA microbial diversity. *The ISME Journal* 3:1365–1373. DOI: 10.1038/ismej.2009.89.

Hyatt D., Chen GL., LoCascio PF., Land ML., Larimer FW., Hauser LJ. 2010. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11. DOI: 10.1186/1471-2105-11-119.

Igarashi K., Kashiwagi K. 1999. Polyamine transport in bacteria and yeast. *Biochem. J.* 344:633–642.

Jorgenson NO., Tranvik LJ., Edling H., Graneli W., Lindell M. 1998. Effects of sunlight on occurrence and bacterial turnover of specific carbon and nitrogen compounds in lake water. *FEMS Microbiology Ecology* 25:217–227.

Kalyuzhnaya MG., Beck DAC., Vorobev A., Smalley N., Kunkel DD., Lidstrom ME., Chistoserdova L. 2011. 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* 62:106–111. DOI: 10.1099/ijs.0.029165-0.

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. *European Journal of Biochemistry* 269:1926–1931. DOI: 10.1046/j.1432-1327.2002.02849.x.

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. DOI: 10.7717/peerj.1165.

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. *Aquatic Microbial Ecology* 68:267–272. DOI: 10.3354/ame01620.

Latypova E., Yang S., Wang Y., 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. *Molecular Microbiology* 75:426–439. DOI: 10.1111/j.1365-2958.2009.06989.x.

Li H., Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595. DOI: 10.1093/bioinformatics/btp698.

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* 2:1–15. DOI: e00169-17.

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* 1:1–6. DOI: 10.1186/2047-217X-1-18.

Magooc T., Salzberg SL. 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. DOI: 10.1093/bioinformatics/btr507.

Markowitz VM., Chen IMA., 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 Research* 40:115–122. DOI: 10.1093/nar/gkr1044.

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. *The ISME Journal* 6:113–123. DOI: 10.1038/ismej.2011.84.

Mou X., Vila-Costa M., Sun S., Zhao W., Sharma S., Moran MA. 2011. Metatranscriptomic signature of exogenous polyamine utilization by coastal bacterioplankton. *Environmental Microbiology* 3:798–806. DOI: 10.1111/j.1758-2229.2011.00289.x.

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 Research* 25:1043–1055.

Paver SF., Kent AD. 2017. Temporal Patterns in Glycolate-Utilizing Bacterial Community Composition Correlate with Phytoplankton Population Dynamics in Humic Lakes. *Microbial Ecology* 60:406–418. DOI: 10.1007/S00248-0.

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 - Molecular Cell Research* 1853:1350–1369. DOI: 10.1016/j.bbamcr.2014.11.021.

Peura S., Eiler A., Bertilsson S., Nyka H., Tiirola M., Jones RI. 2012. Distinct and diverse anaerobic bacterial communities in boreal lakes dominated by candidate division OD1. *The ISME Journal* 6:1640–1652. DOI: 10.1038/ismej.2012.21.

Peura S., Sinclair L., Bertilsson S., Eiler A. 2015. Metagenomic insights into strategies of aerobic and anaerobic carbon and nitrogen transformation in boreal lakes. *Scientific Reports* 5:12102. DOI: 10.1038/srep12102.

Pomeroy LR., Wiebe WJ. 1988. Energetics of microbial food webs. *Hydrobiologia* 159:7–18. DOI: 10.1007/BF00007363.

Price MN., Dehal PS., Arkin AP. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLOS ONE* 5. DOI: 10.1371/journal.pone.0009490.

Ramachandran A., Walsh DA. 2015. Investigation of XoxF methanol dehydrogenases reveals new methylotrophic bacteria in pelagic marine and freshwater ecosystems. *FEMS Microbiology Ecology* 91. DOI: 10.1093/femsec/fiv105.

Ramanan R., Kim B-H., Cho D-H., Oh H-M., Kim H-S. 2015. Algae–bacteria interactions: evolution, ecology and emerging applications. *Biotechnology Advances*. DOI: 10.1016/j.biotechadv.2015.12.003.

Remsen CC., Carpenter EJ., Schroeder BW. 1972. Competition for Urea among Estuarine Microorganisms. *Ecological Society of America* 53:921–926.

Rinke C., Schwientek P., Sczyrba A., Ivanova NN., Anderson IJ., Cheng J-F., Darling AE., Malfatti S., Swan BK., Gies E a., Dodsworth J a., Hedlund BP., Tsiamis G., Sievert SM., Liu W-T., Eisen J a., 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. DOI: 10.1038/nature12352.

Roux S., Chan LK., Egan R., Malmstrom RR., McMahon KD., Sullivan MB. 2017. Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. *Nature Communications* 8. DOI: 10.1038/s41467-017-01086-2.

Salcher MM., Neuenschwander SM., Posch T., Pernthaler J. 2015. The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. *The ISME Journal* 9:2442–2453. DOI: 10.1038/ismej.2015.55.

Salcher MM., Posch T., Pernthaler J. 2013. In situ substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *Isme J* 7:896–907. DOI: 10.1038/ismej.2012.162.

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. *Applied and Environmental Microbiology* 75:7537–7541. DOI: 10.1128/AEM.01541-09.

Segata N., Waldron L., Ballarini A., Narasimhan V., Jousson O., Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods* 9:811–4. DOI: 10.1038/nmeth.2066.

Šimek K., Kasalický V., Jezbera J., Jezberová J., Hejzlar J., Hahn MW. 2010. Broad habitat range of the phylogenetically narrow R-BT065 cluster, representing a core group of the betaproteobacterial genus limnohabitans. *Applied and Environmental Microbiology* 76:631–639. DOI: 10.1128/AEM.02203-09.

Sommer DD., Delcher AL., Salzberg SL., Pop M. 2007. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* 8:64. DOI: 10.1186/1471-2105-8-64.

Tang KH., Blankenship RE. 2010. Both forward and reverse TCA cycles operate in green sulfur bacteria. *Journal of Biological Chemistry* 285:35848–35854. DOI: 10.1074/jbc.M110.157834.

Varghese NJ., Mukherjee S., Ivanova N., Konstantinidis KT., Mavrommatis K., Kyrpides NC., Pati A. 2015. Microbial species delineation using whole genome sequences. *Nucleic Acids Research* 43:gkv657-. DOI: 10.1093/nar/gkv657.

Walsh JR., Carpenter SR., Vander Zanden MJ. 2016. Invasive species triggers a massive loss of ecosystem services through a trophic cascade. *Proceedings of the National Academy of Sciences* 113:4081–4085. DOI: 10.1073/pnas.1600366113.

Williamson CE., Dodds W., Kratz TK., Palmer MA. 2008. Lakes and streams as sentinels of environmental change in terrestrial and atmospheric processes. *Frontiers in Ecology and the Environment* 6:247–254. DOI: 10.1890/070140.

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 Research* 40:445–451. DOI: 10.1093/nar/gks479.