# 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). Approximately half of the carbon received by freshwater ecosystems from the terrestrial landscape is emitted as carbon dioxide (0.2 Pg C/year) or stored (0.8 Pg C/year) (Cole et al., 2007). Similarly, 20% of global denitrification is estimated to occur in freshwater, roughly equivalent to the amount of denitrification taking place in soils (22%) and about a third of the amount occurring in oceans (58%) (Seitzinger et al., 2006). Because of these globally relevant contributions to carbon and nutrient cycling, understanding how these elements are processed in freshwater through methods such as models and budgets is an important area of research.

Much of this freshwater biogeochemical cycling is performed by microbial communities. However, the categories of carbon and nutrients often included in models of freshwater biogeochemical cycling are too broad to be used to incorporate microbial data. For example, carbon compounds are often classified as labile and recalcitrant (Guillemette & del Giorgio, 2011), or autochthonous and allochthonous (Jonsson et al., 2001). While some work has been done on microbial responses to these carbon categories (Eiler et al., 2003; Kritzberg et al., 2004), using such broad categorizations masks much of the complexity of microbial substrate use. Similarly, elements such as nitrogen and phosphorus are often classified as simply organic or inorganic. Incorporating microbially-mediated transformations of specific compounds in freshwater would significantly improve the accuracy and predictive power of biogeochemical cycling models.

However, linking microbial taxa to specific functions is a challenging task. Previous research has investigated substrate use by freshwater taxa using cultured isolates or microscopy techniques (Hahn et al., 2012; Salcher, Posch & Pernthaler, 2013b). While this work is highly informative, it cannot be scaled to investigate many community members simultaneously. Genomics techniques can be applied at the community level, and although any functional predictions from genomic data are merely predictions, they are still powerful tools for analyzing microbial communities. Sequencing data has previously been employed to great effect to analyze the distribution of functional marker genes in freshwater (Ramachandran & Walsh, 2015; Peura et al., 2015) and to predict metabolic potential in freshwater taxa (Hamilton et al., 2017).

In this research, we combine insights from both genes and genomes in multiple freshwater metagenomic time series to link function to taxonomy on at the community level. Our metagenomic time series include multiple years of sampling for microbial DNA from two lakes in Wisconsin, USA: Lake Mendota, a large eutrophic lake, and Trout Bog, a small humic lake. Lake Mendota and Trout Bog are ideal sites for comparative time series metagenomics because of their history of extensive environmental sampling by the North Temperate Lakes - Long Term Ecological Research program (NTL-LTER, <http://lter.limnology.wisc.edu>) and their contrasting limnological attributes (Table 1, Table S1). We analyze both predicted pathways in metagenome-assembled genomes (MAGs) and the distributions of functional marker genes to provide a comprehensive overview of microbially-mediated biogeochemical cycling in two contrasting freshwater lakes.

Throughout this paper, we highlight several functional categories with particularly interesting results. We discuss differences in the identity and diversity of potential nitrogen fixing bacteria in Trout Bog vs. Lake Mendota, as well as the high prevalence of genes related to polyamines, proposed to be an important component of the dissolved organic nitrogen pool. We observed that assimilatory sulfate reduction pathways were encoded more frequently than dissimilatory sulfate reduction pathways, in contrast to what is thought to be the case in marine systems. We split the broader category of primary production into different types of phototrophy, including photosynthesis performed by Cyanobacteria, green sulfur bacteria, and aerobic anoxygenic phototrophs, and analyzed their associated carbon fixation pathways (when present). Using annotations of carbohydrate-active enzymes, we compared the potential for complex carbon degradation and found significant differences in the coding density and diversity of these encoded enzymes between lakes. To compare more basic properties of freshwater microbes, we assessed differences between lakes in aspects of more central microbial metabolisms such as hydrogen metabolism, oxidative phosphorylation, methylotrophy, and degradation of low molecular weight carbon. Finally, we show how trends over time in the abundances of both nitrogen fixation marker genes and Cyanobacteria MAGs likely encoding nitrogen fixation were highly correlated, demonstrating how genomic data can reveal dynamics in both functions and taxa.

We anticipate that this dataset will be a valuable community resource for other freshwater microbial ecologists to mine and incorporate into comparative studies across lakes around the world. As such, all data is publicly available at < mcmahon lab github link>. The results of this study can be used to guide efforts to build microbially-resolved models of freshwater carbon and nitrogen cycles with better predictive power.