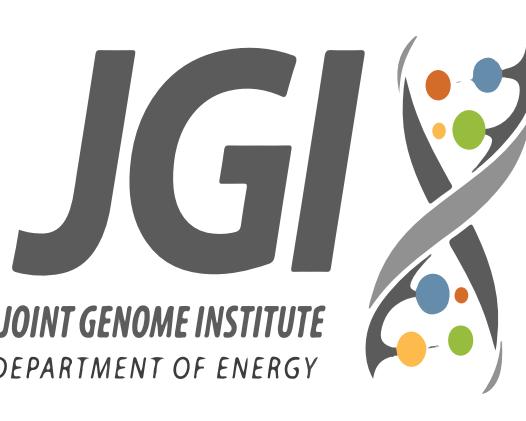




GRFP

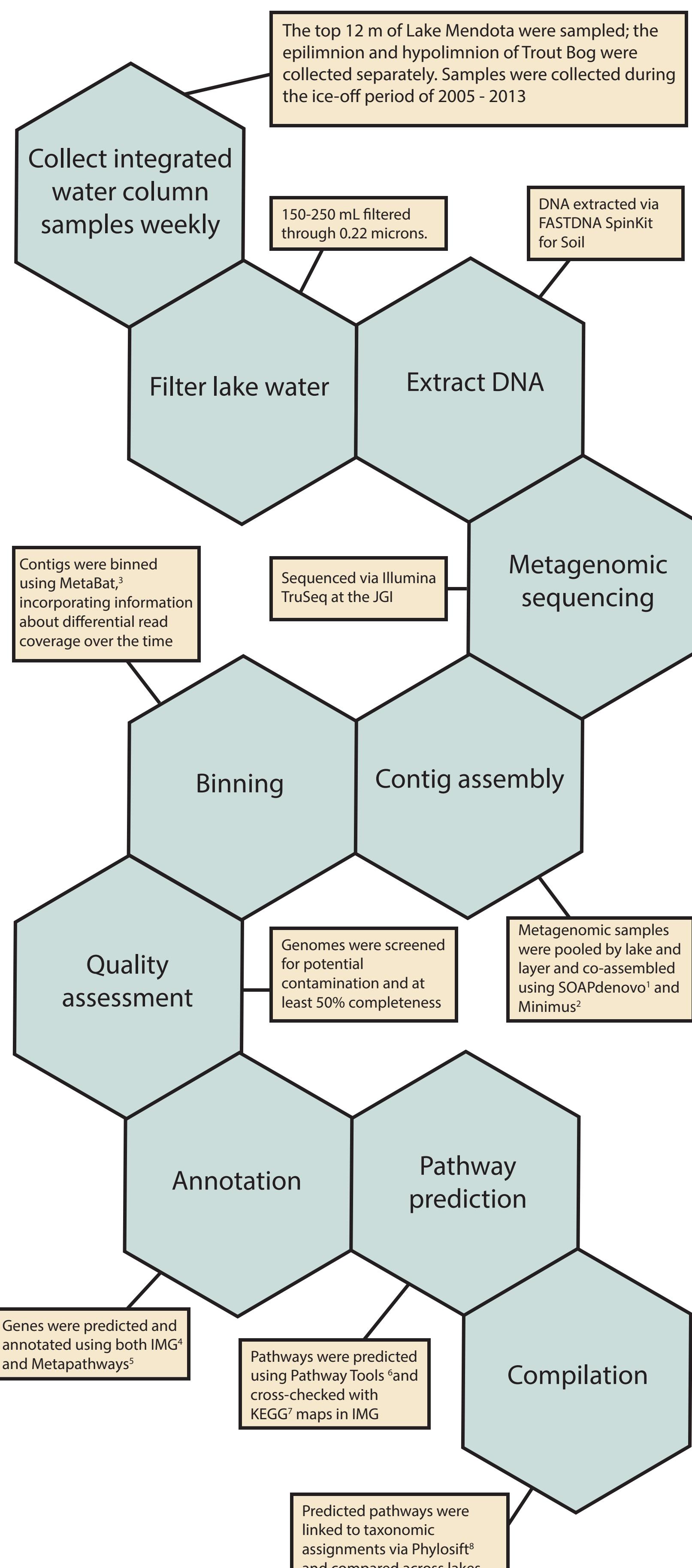
# Functional Prediction of Freshwater Bacterial Genomes

Alexandra Linz<sup>1</sup>, Shaomei He<sup>1</sup>, Samuel Schmitz<sup>1</sup>, Sarah Stevens<sup>1</sup>, Rex Malmstrom<sup>2</sup>, Katherine McMahon<sup>1</sup><sup>1</sup>University of Wisconsin - Madison, <sup>2</sup>Department of Energy - Joint Genome Institute

## Background

Bacteria form the base of freshwater ecosystems because of their importance in nutrient cycling. We used genomes assembled from metagenomes to predict metabolisms of bacterial populations in two lakes, Lake Mendota and Trout Bog, in Wisconsin, USA. We know that environmental parameters and bacterial community composition differ between these lakes, but do not know how the functions of these communities compare. Analyzing potential functions at the population level allows us to integrate phylogeny and metabolic information from gene annotations. By comparing the predicted pathways of bacterial communities in two lakes with different nutrient regimes, we can link function at the population scale to the ecosystem scale.

## Methods



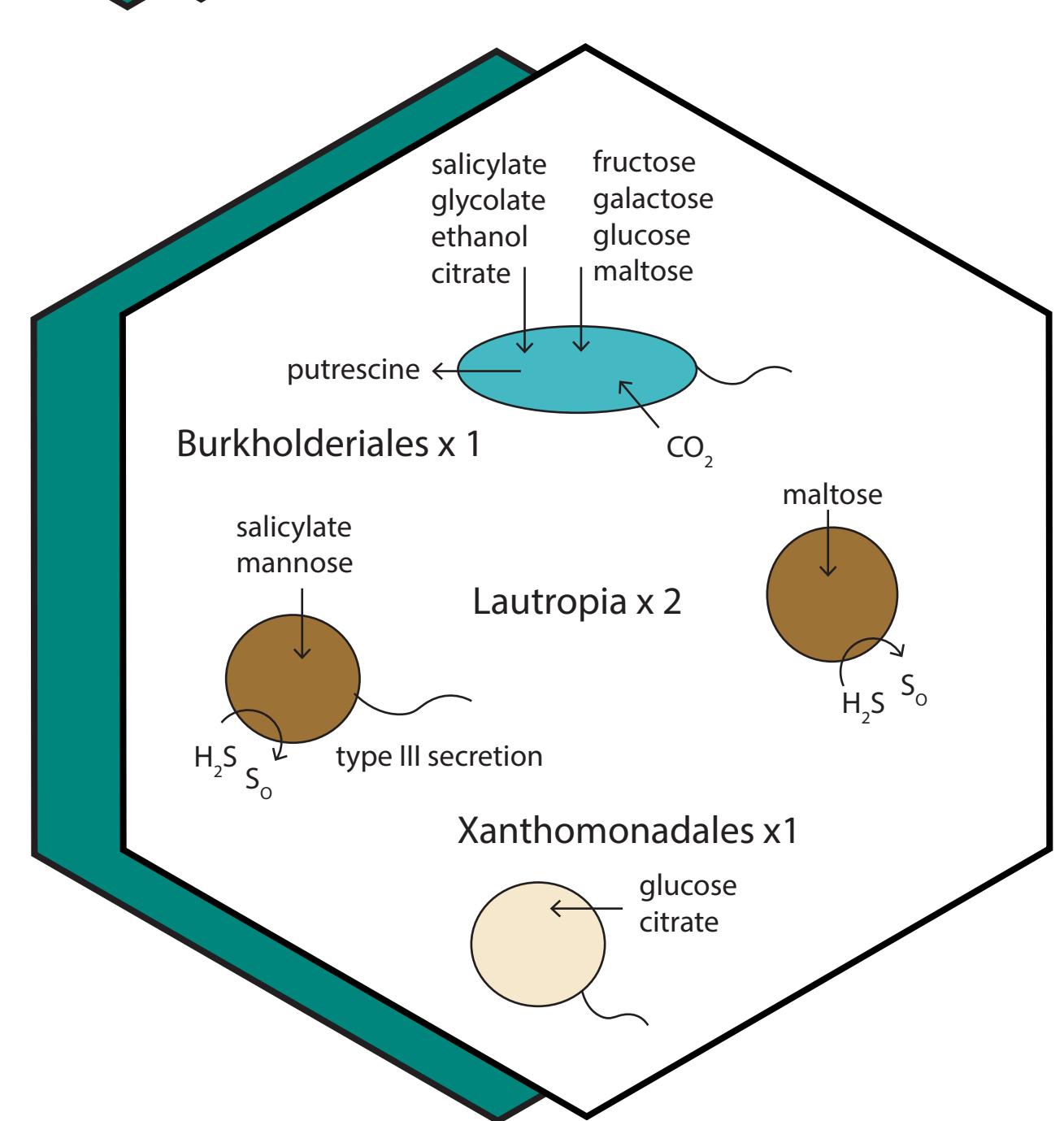
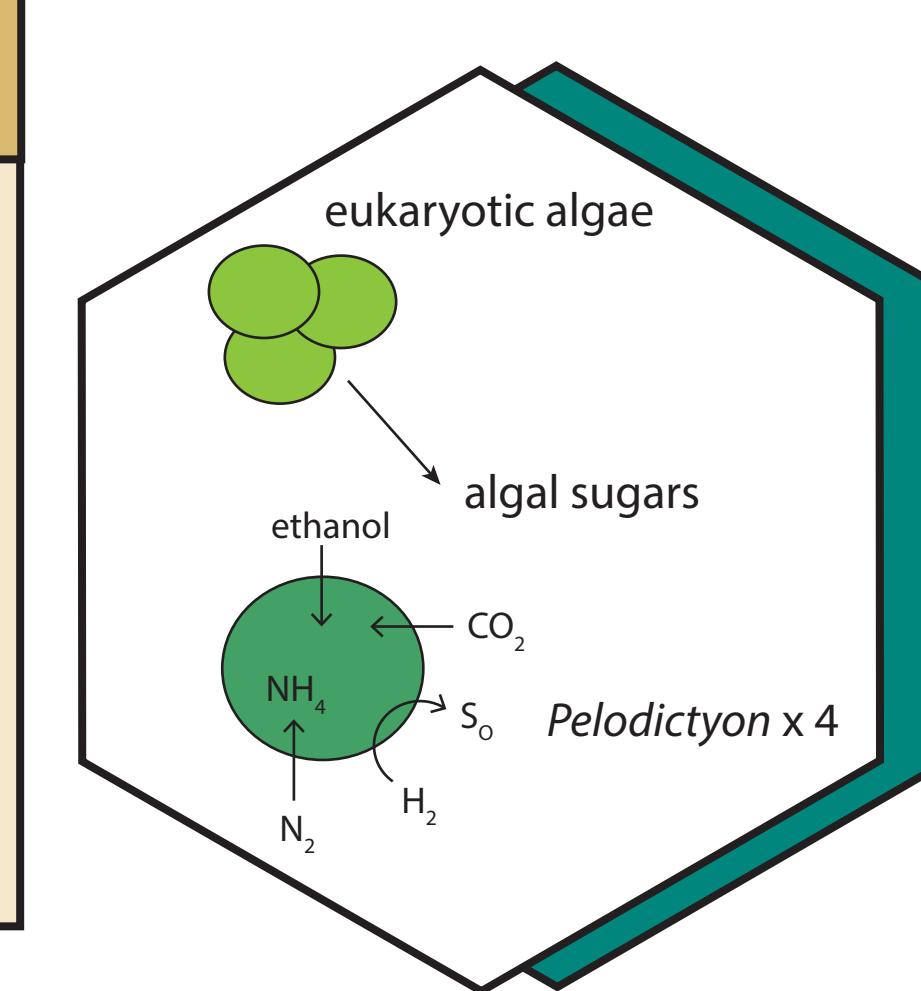
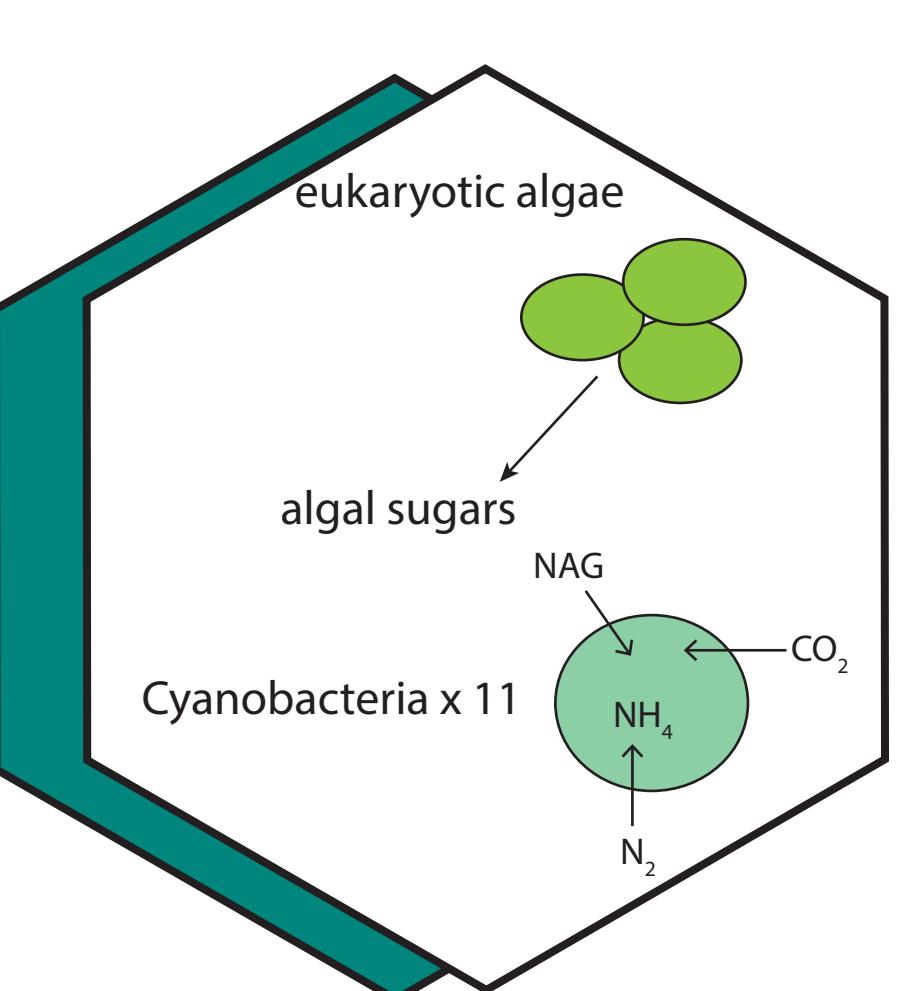
## References

- Xie, Yinlong, et al. "SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads." *Bioinformatics* 30.12 (2014): 1660-1666.
- Sommer, Daniel D., et al. "Minimus: a fast, lightweight genome assembler." *BMC bioinformatics* 8.1 (2007): 64.
- Kang, Dongwan, et al. "MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities." *PeerJ* 3 (2015): e1165.
- Markowitz, Victor M., et al. "The integrated microbial genomes (IMG) system." *Nucleic acids research* 34.suppl. 1 (2006): D344-D348.
- Konwar, Kishori M., et al. "MetaPathways: a modular pipeline for constructing pathway/genome databases from environmental sequence information." *BMC bioinformatics* 14.1 (2013): 202.
- Karp, Peter D., et al. "Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology." *Briefings in bioinformatics* 11.1 (2009): 40-79.
- Kanehisa, Minoru, and Susumu Goto. "KEGG: Kyoto encyclopedia of genes and genomes." *Nucleic acids research* 28.1 (2000): 27-30.
- Darling, Aaron E., et al. "PhyloSift: phylogenetic analysis of genomes and metagenomes." *PeerJ* 2 (2014): e243.
- Repetto, Daniel J., et al. "Chemical characterization of high molecular weight dissolved organic matter in fresh and marine waters." *Geochimica et Cosmochimica Acta* 66.6 (2002): 955-962.

## Primary Production

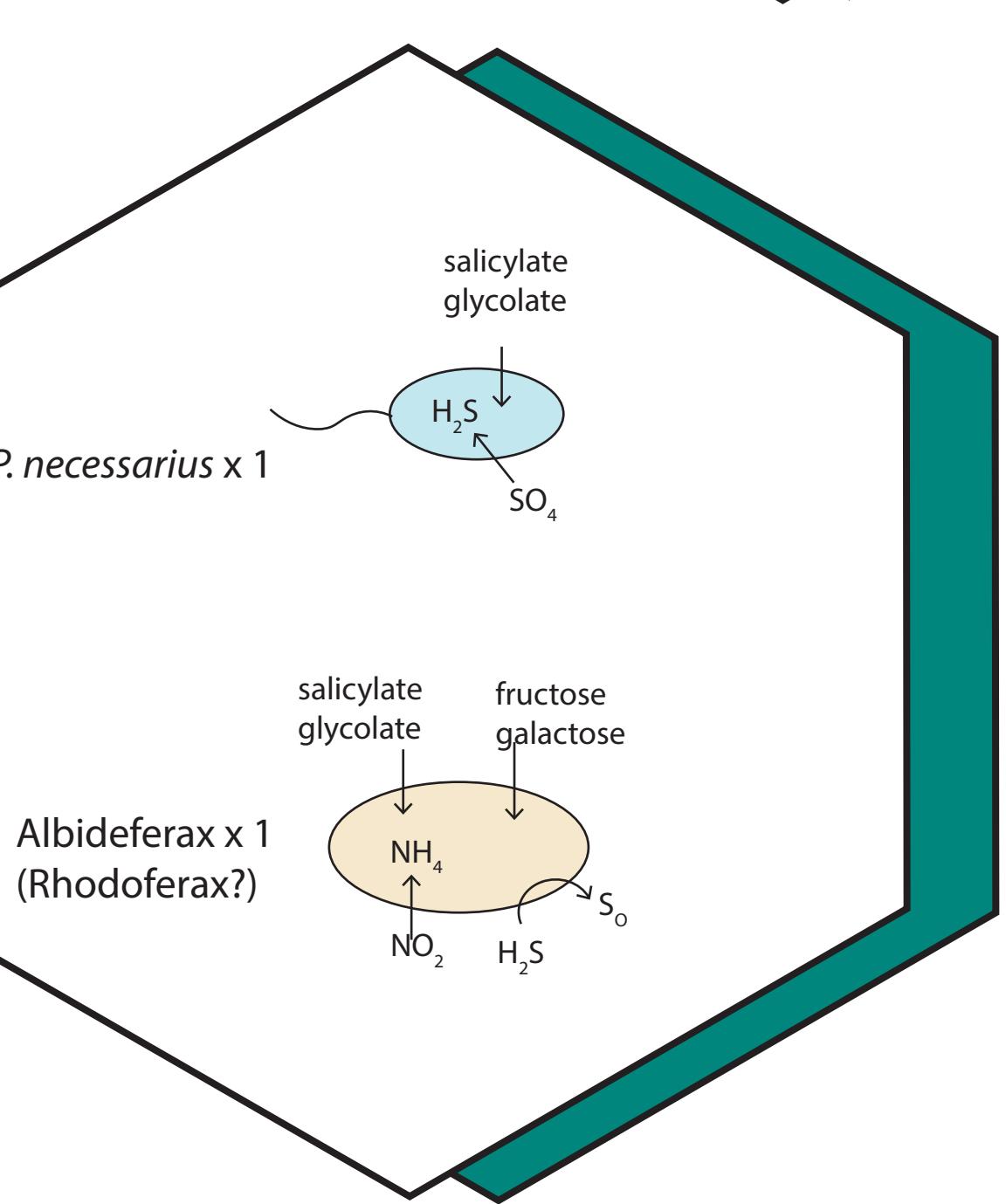
Photoautotrophy is an important ecosystem function in both Lake Mendota and Trout Bog. In Lake Mendota, primary production is performed by Cyanobacteria in the presence of oxygen. These bacteria fix carbon via the Calvin-Benson-Bassham pathway. In Trout Bog, *Pelodictyon* of the phylum Chlorobi are the main primary producers. *Pelodictyon* use hydrogen sulfide instead of water as the electron donor for photosynthesis, and fix carbon via the reductive TCA cycle. While photoautotrophy is a key function in both lakes, the differences in the pathways used reflect the differences in available compounds in each ecosystem.

Eukaryotic algae, while not included in this analysis, are present in both lakes. They produce sugars such as glucose, galactose, mannose, xylose, and arabinose<sup>9</sup>. Pathways for the degradation of algal sugars are common in both lakes.



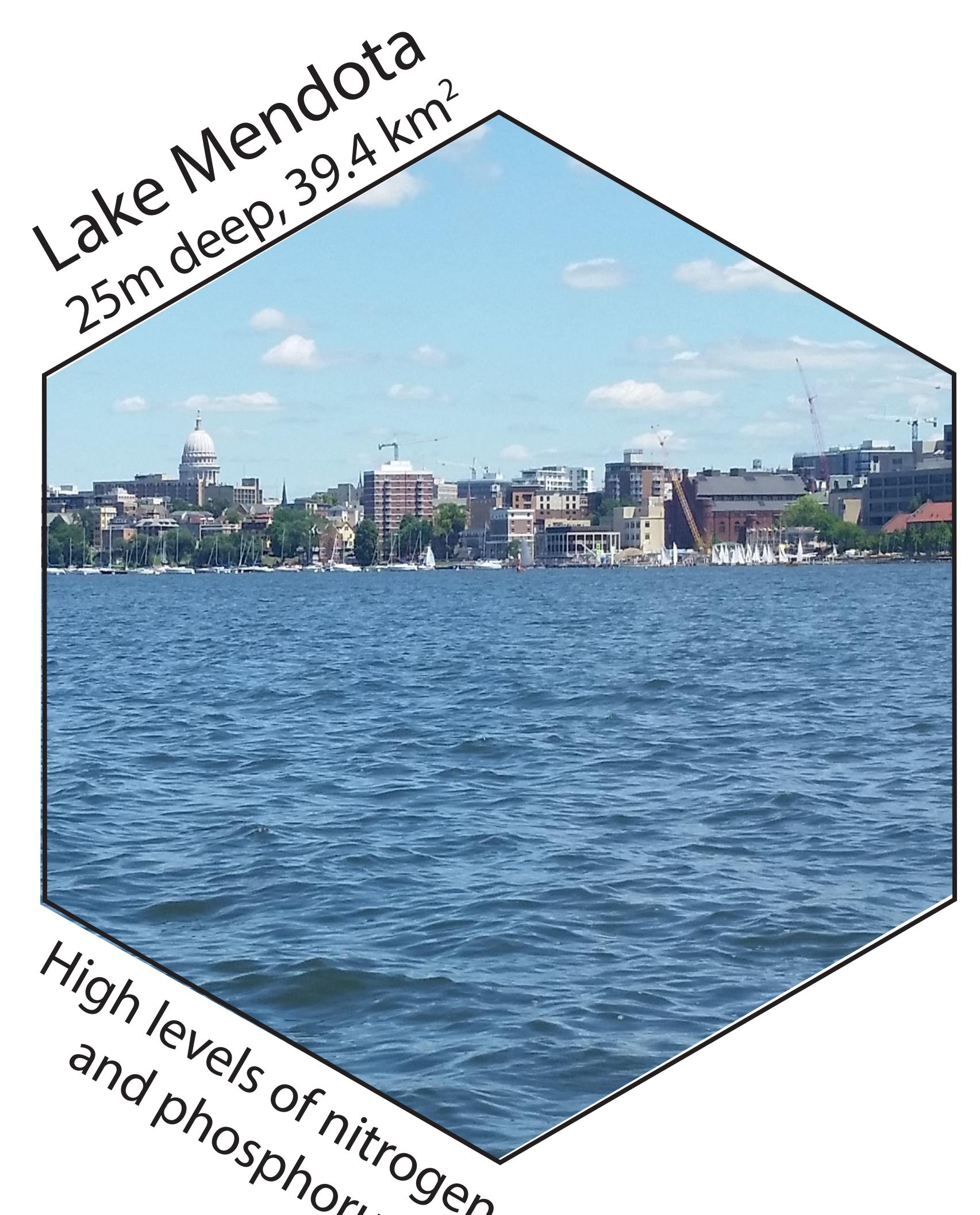
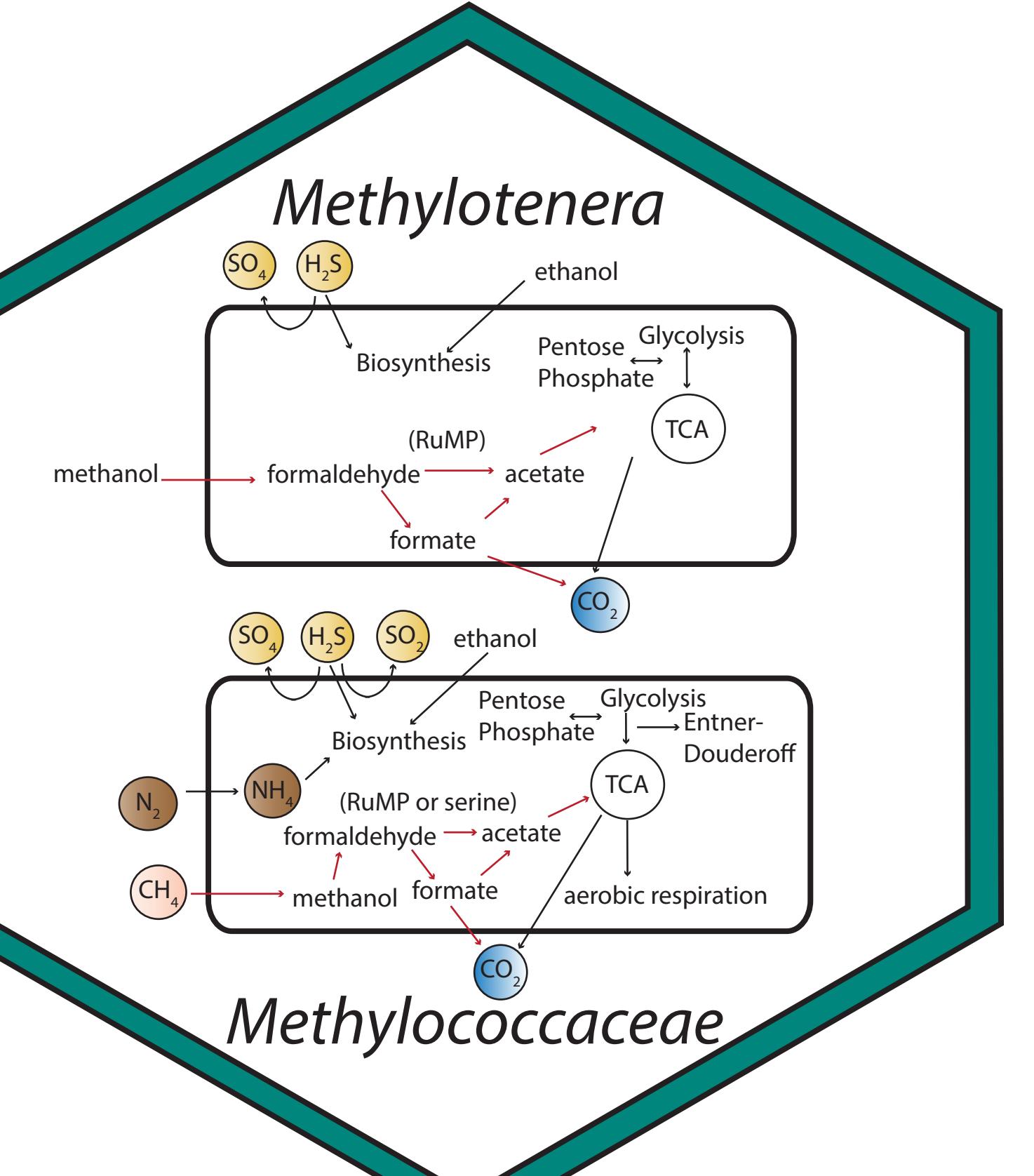
## Photoheterotropy

Both lakes contained bacteria that used photosynthesis to fuel heterotrophy rather than carbon fixation. These bacteria belonged mainly to Betaproteobacteria, with one exception in Gammaproteobacteria, and contained pathways for the degradation of both terrestrial and aquatic carbon.



## Methylotrophy

Methylotrophy, the degradation on one carbon compounds such as methane or methanol, is performed by the same bacteria using the same pathways in both lakes.



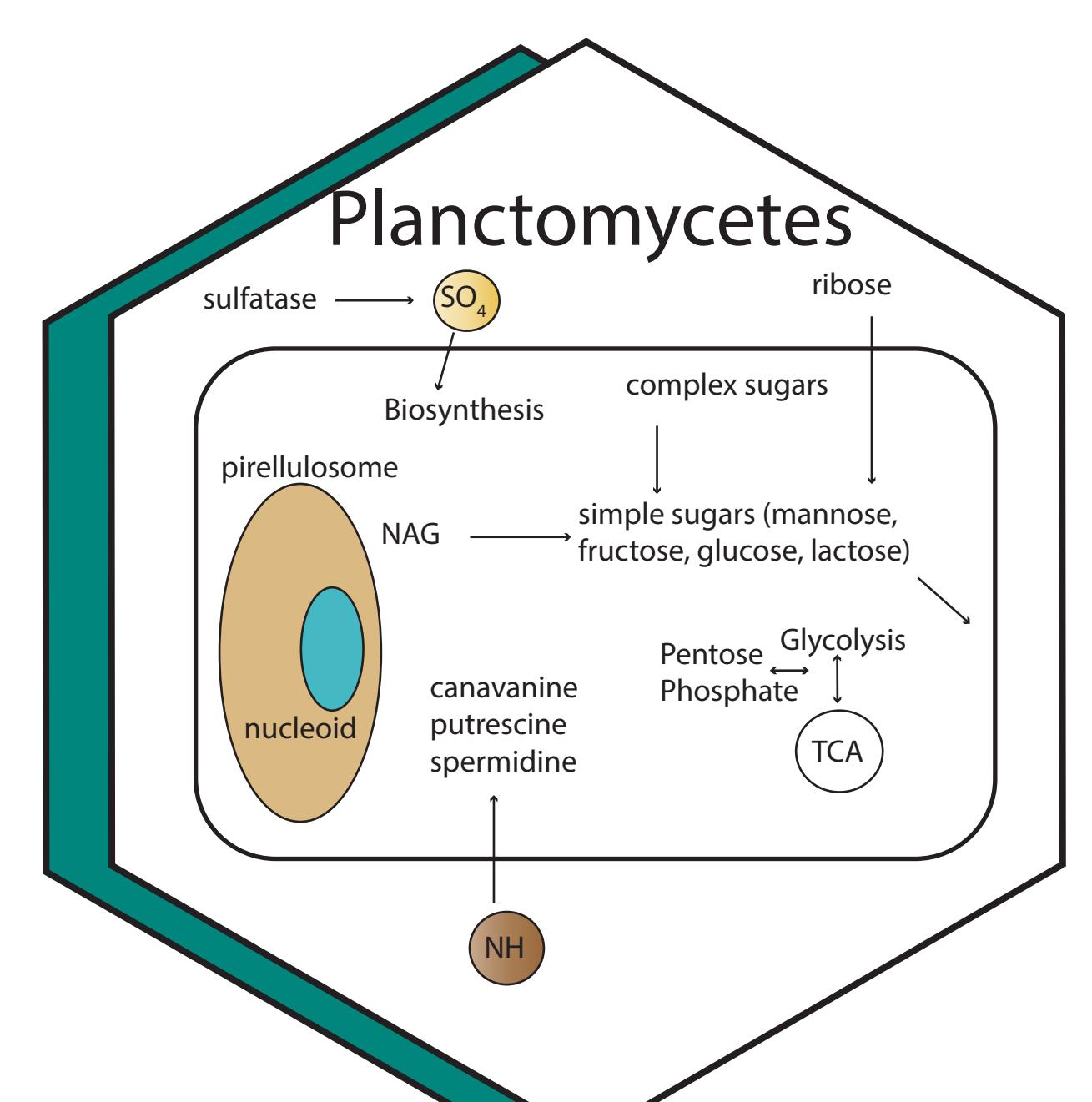
**Lake Mendota**  
25m deep, 39.4 km<sup>2</sup>

High levels of nitrogen and phosphorus



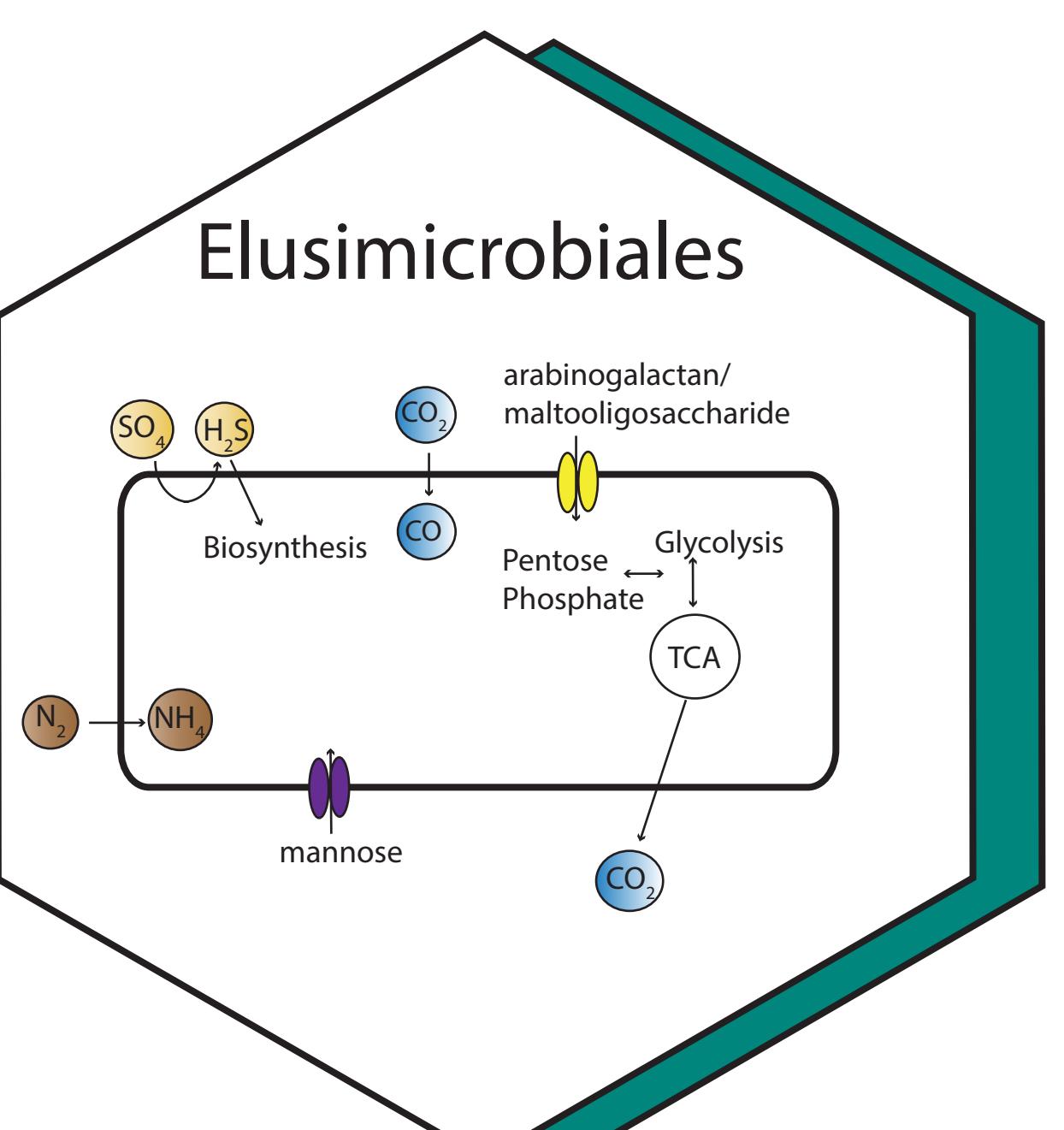
**Trout Bog**  
7m deep, 0.01 km<sup>2</sup>

High carbon, low nitrogen and phosphorus



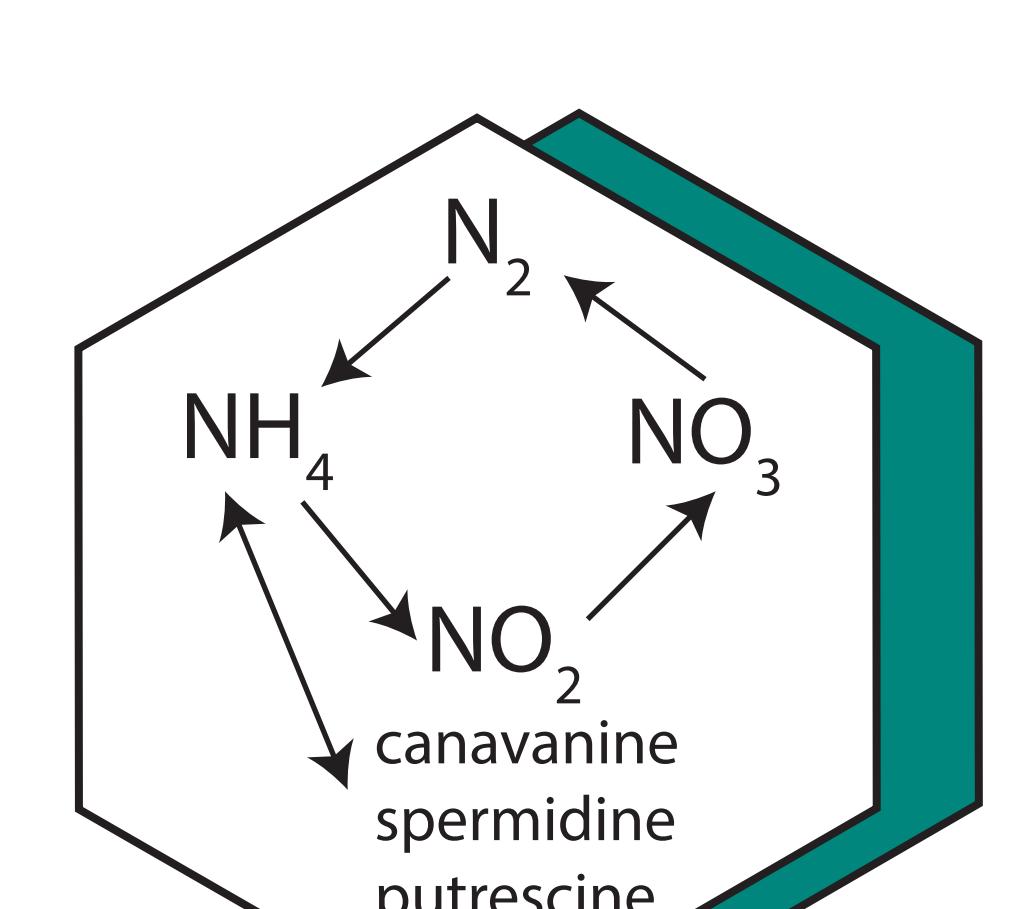
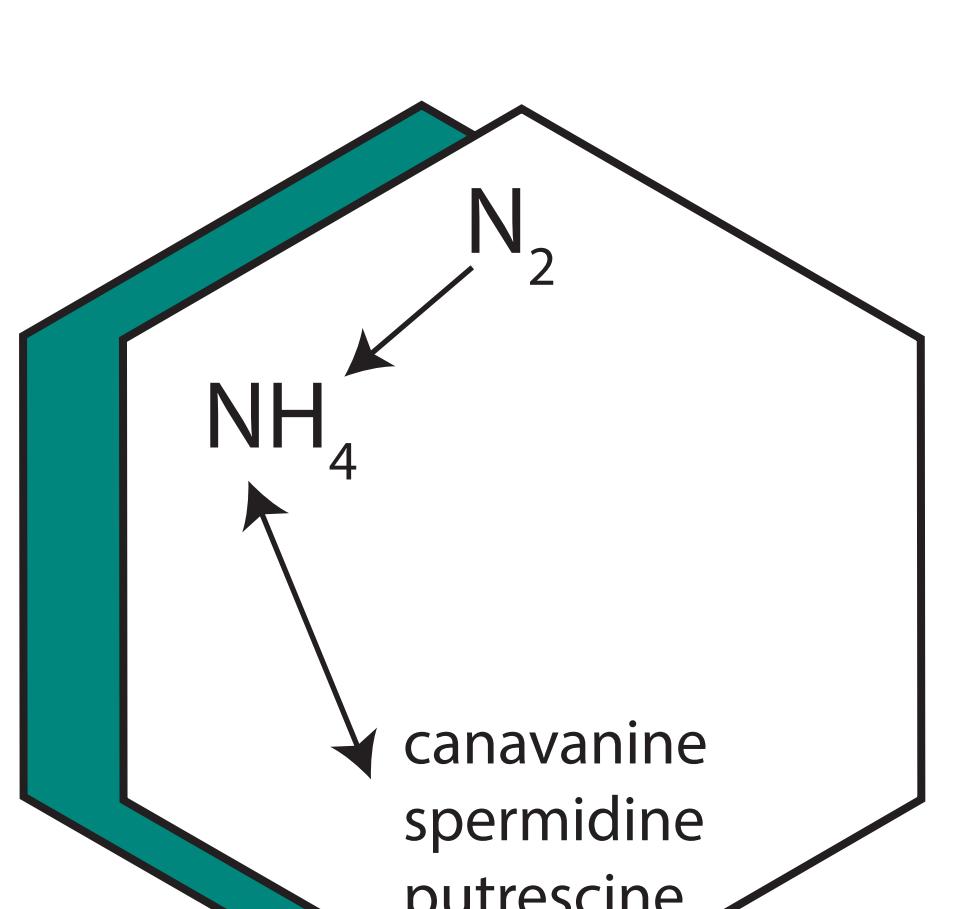
## Unusual Bacteria

Genomes from bacterial groups with few cultured representatives, particularly from freshwater, can lend insight into their metabolisms. We recovered an Elusimicrobiales genome from Trout Bog capable of sulfate reduction, nitrogen fixation, and degradation of sugars. Genomes from Planctomycetes contained genes encoding the biosynthesis of canavanine, spermidine, and putrescine, as well as sulfatases and glucosidase hydrolases, often used to degrade polysaccharides.

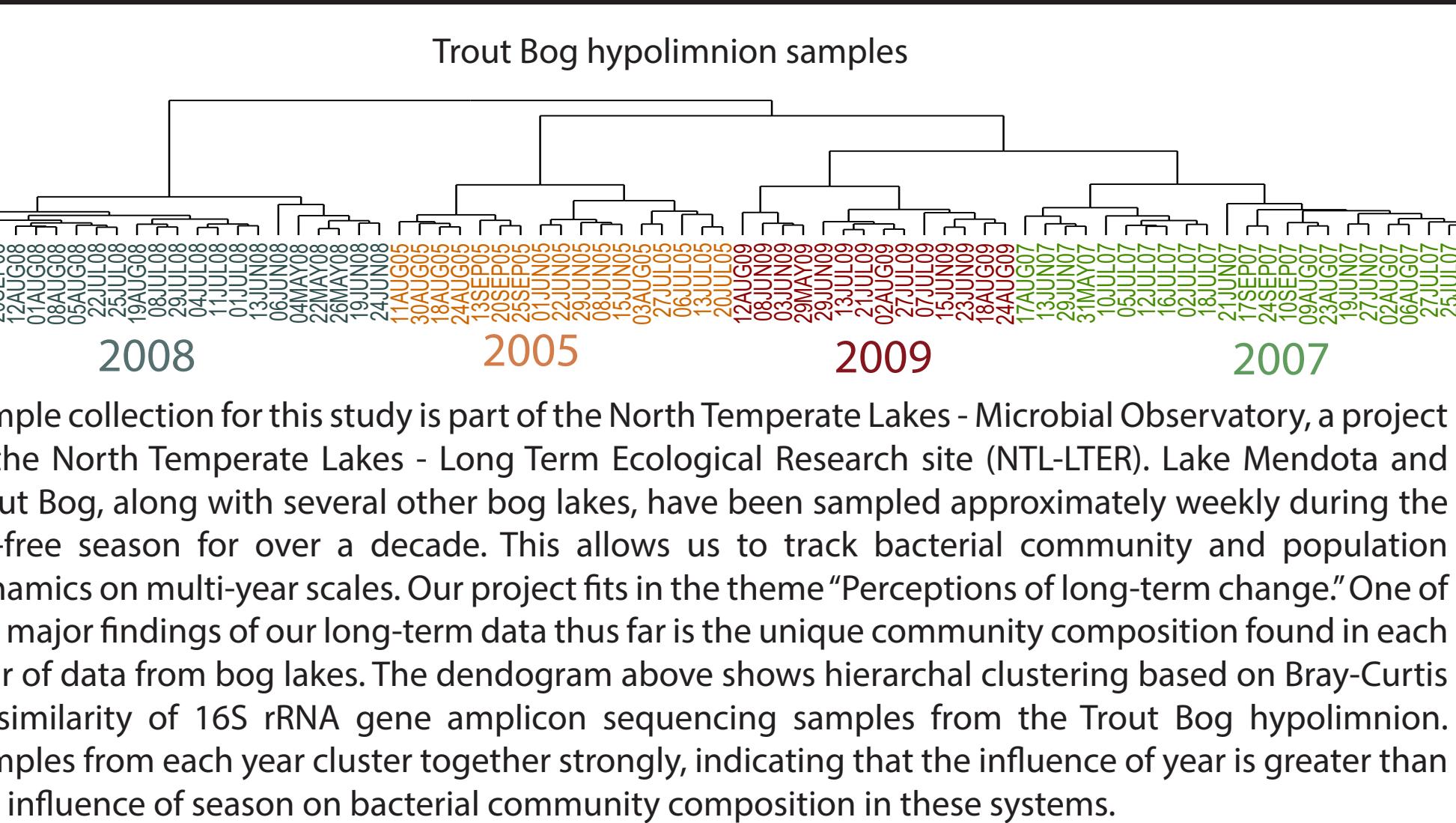


## Nitrogen Cycling

Trout Bog is a nitrogen-limited system, while Lake Mendota receives high levels of nitrogen from agricultural runoff. This is reflected in the nitrogen cycling pathways found in genomes from each lake. Many phylogenetically diverse bacteria contain genes for nitrogen fixation in Trout Bog, while only Cyanobacteria fix nitrogen in Lake Mendota. Few genes encoding nitrification or denitrification are found in Lake Mendota compared to Trout Bog. However, non-proteinogenic amino acids and polyamines appear to be important intermediates in nitrogen cycling in both ecosystems. Pathways for the biosynthesis and degradation of canavanine, spermidine, and putrescine were identified in both Lake Mendota and Trout Bog.

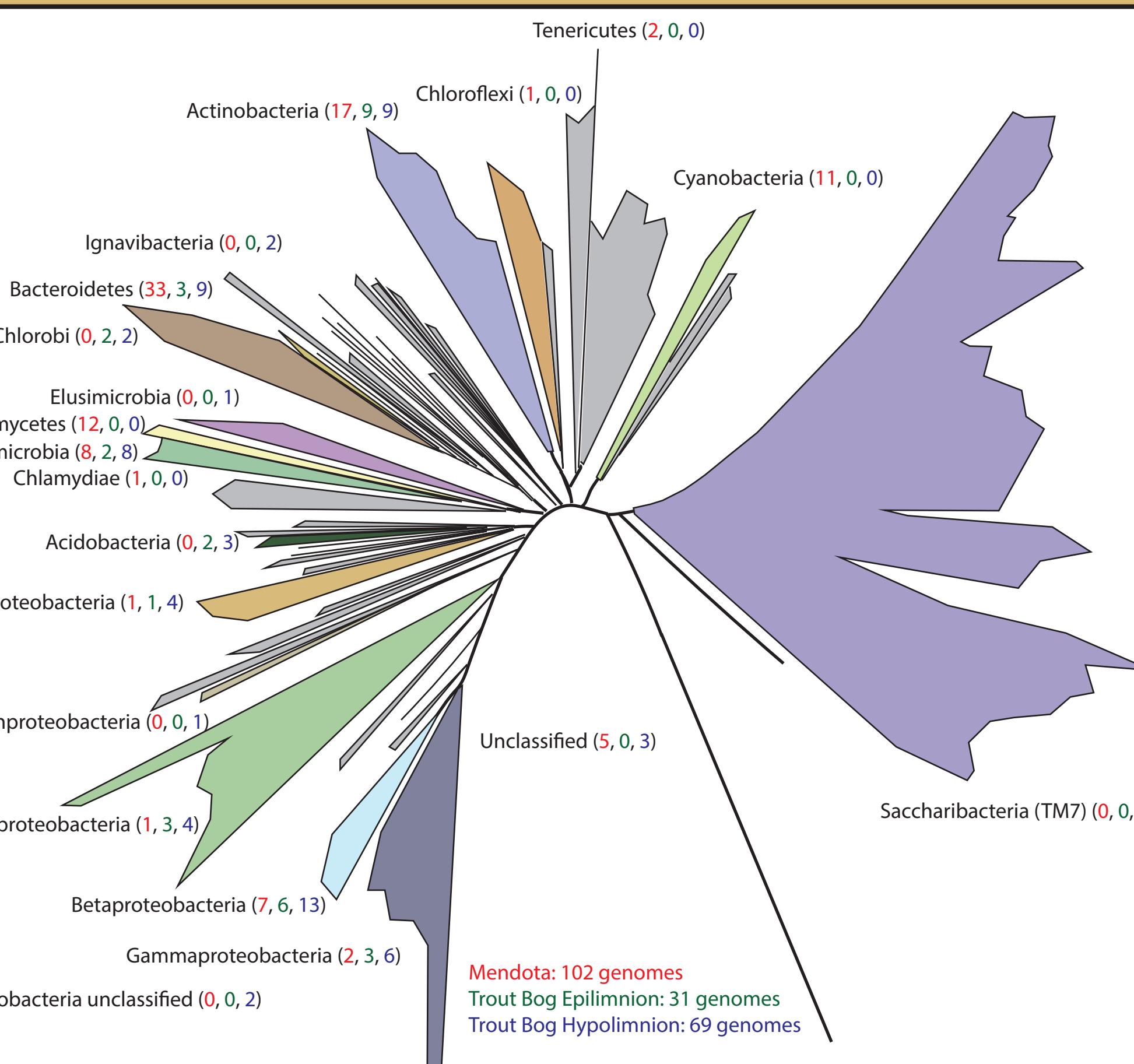


## Long Term Ecological Research

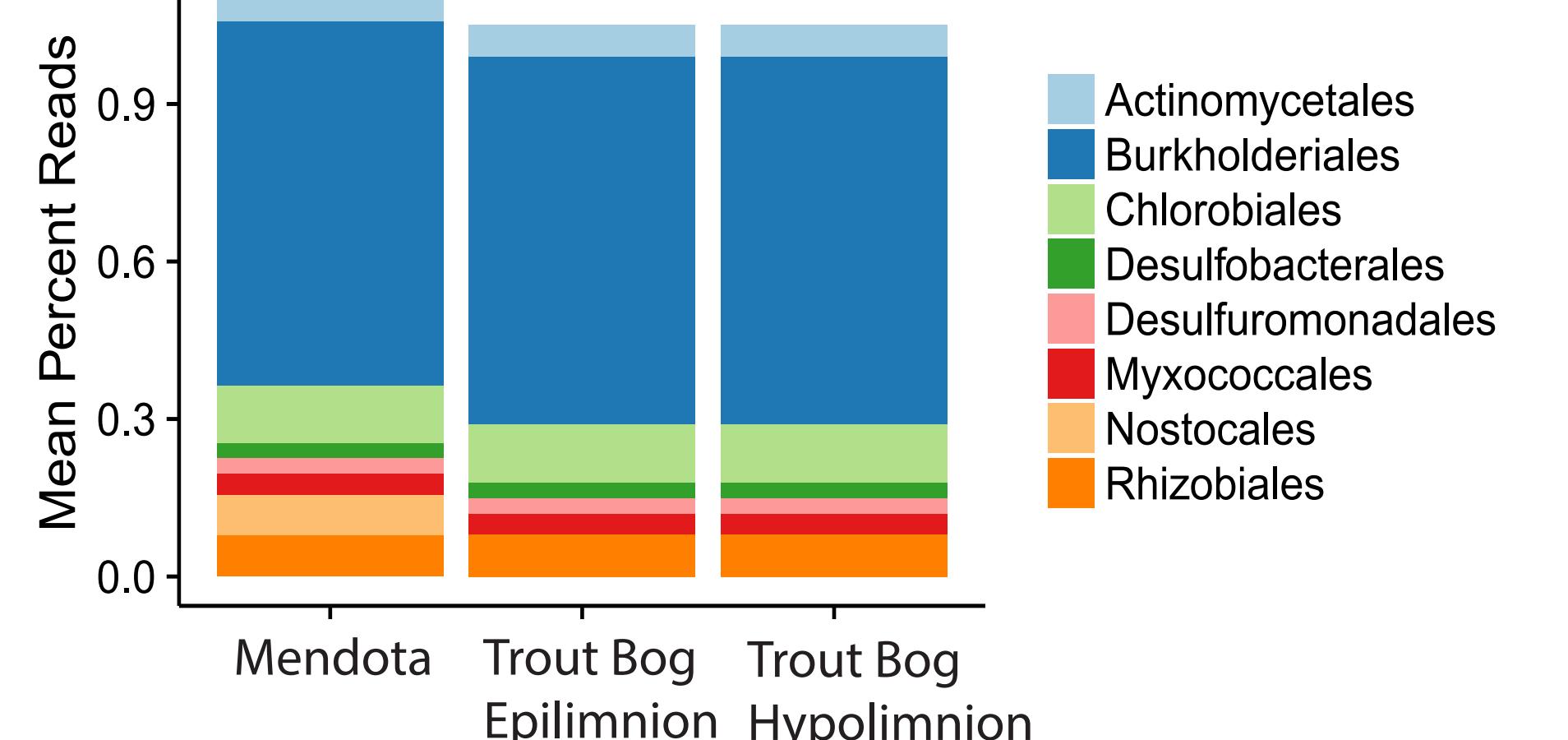


Having multiple years of sampling benefited this project in two ways. First, it allowed better genome reconstruction because read coverage over time could be used to group DNA sequencings with similar abundance patterns. Second, the high amount of inter-annual variability in community composition means that including multiple years of data increased the observed diversity of genomes recovered.

## Genome Information



## Read Classifications



**Classification of metagenomic reads.** Metagenomic reads from each lake were classified using kraken. Results are largely consistent with the classifications of recovered genomes; for example, Lake Mendota contains reads classified as Cyanobacterial class Nostocales, while Trout Bog does not. Reads classified as Chlorobiales are likely instead Bacteroidetes, as this closely related group was found more frequently in genomes, and Chlorobiales is not typically found in Lake Mendota.

## Conclusions and Future Work

### Take home points:

- Broad level functions (such as primary production) are shared between lakes, but are performed by different taxa using different pathways
- Nitrogen cycling reactions present reflect nutrient concentrations in each lake, but polyamines and non-proteinogenic amino acids are pools of dissolved organic nitrogen in both systems.
- Analysis of predicted pathways can suggest functions for bacteria with few cultured relatives

### Future directions:

- More in depth analyses of rhodopsins and polysaccharide degradation
- Gene expression in Actinobacteria
- Diel trends in gene expression
- Continued weekly sampling