**Specific Aims**

Goal: Learning about the ecology and evolution of microbes in lakes using genomics, including how populations change over time in lakes, metabolic potential of the lake microbes, and genomic features of the lake microbes.

**AIM 1:**

Bacterial evolution in lakes has not previously been well studied due the lack of cultivability for many microbes and inability to track natural populations on a whole genome scale. Efforts have been made to compare natural populations using whole genomes from cultivated organisms from a single environment. There have also been studies that have mapped metagenomes to genomes from cultivated organisms. However, neither of these methods considers the uncultivated majority or attempts to look at the bacterial populations in a single ecosystem on a broader scale. In lakes, many of the abundant microbes have yet to be cultivated and isolated, such as the acI clade of Actinobacteria. By looking at only the cultivated bacteria, microbes performing important ecosystem functions may be missed. Using our metagenomic time series, which spans 3 years and two layers in Trout Bog and 5 years in Lake Mendota, we can use coverage based binning and sequence features to retrieve reference genomes, which are directly relevant to our data set and not necessarily cultivatable. We also have single amplified genomes, SAGs, from Lake Mendota, which represent a variety of major freshwater lineages. We can map metagenomic reads back onto these genomes from metagenomes, GFMs, and the SAGs and investigate the population dynamics over time. This approach is relatively novel since using coverage based binning to retrieve genomes from metagenomes is a recently developed technique. We also have an exceptional dataset with our time series of the two lakes over several years. I have worked with collaborators at the Joint Genome Institute (JGI) on the evolutionary dynamics, specifically single nucleotide polymorphisms (SNPs) and gene gain and loss, in a smaller set of manually binned genomes, for which there is a paper currently submitted.

**To determine the evolutionary dynamics of sequence discrete populations of bacteria in freshwater lakes, we will map reads from metagenomic time series to composite GFMs and SAGs and examine SNP patterns, gene gain and loss, recombination, and abundance patterns.**

*H1.1: As predicted due to their low diversity and low rates of recombination, the LD12 lineage experiences genome-wide sweeps over time.*

*H1.2: The acI lineage, which has much greater diversity in comparison with the LD12 lineage, will undergo gene-specific sweeps over time and has higher rates of recombination.*

*H1.3: Each lineage undergoes gene-specific or genome-wide sweeps when under selective pressure, depend ending on their rates of recombination.*

**AIM 2:**

There is little known about the functions these freshwater bacteria are performing in the environment. Without cultivation, experiments cannot be performed to directly test the uptake of substrates. Microautoradiography and fluorescence *in situ* hybridization (MAR-FISH) can be used to observe which cells are taking up a radio-labeled substrate. In order to do MAR-FISH there must be something known about which organism might take up that substrate to design probes. To look for functional capabilities, we can use the GFMs and SAGs to explore functional potential as indicated by the gene predictions. This will provide lineage specific information about possible substrate uptake in lakes. We can use the evolutionary patterns found above to target functions experiencing selective pressures in the environment. Studies have been done, including some in the McMahon lab, to determine functional capabilities using SAGs(Sarahi’s paper, Trevor’s Paper, Martinez-Garcia VerrucoPaper). We can also target specific groups and their functional capabilities by looking for correlations between specific nutrient conditions and relative abundances determined from mapping metagenomic reads to the genomes. Learning more about the metabolic functions microbes are performing in the environment can help us to better understand the nutrient cycling being carried out.

**To hypothesize which bacterial groups are performing specific metabolic functions in the community, we will characterize the functional potential of bacteria using GFMs and SAGs.**

**AIM 3:**

Genome streamlining has been observed in freshwater and marine settings but little is understood about the features that make this lifestyle successful in aquatic environments. As previously mentioned, freshwater systems have very few reference genomes and many of those previously studied were cultivatable. It has also been suggested that previous difficulty in culturing these bacteria may be due to their streamlined nature(Giovannoni et al., 2005). The genomes from cultivatable bacteria may not be representative of those microbes that are most abundant or are performing important functions in the environment. To learn if streamlined genomes are common among freshwater microbes, we need to analyze the genomes from uncultivated organisms. We propose to characterize the genome features of the same GFMs and SAGs from aims one and two to find features of genome streamlining. This substantial data set provides the opportunity to compare the whole genomes, many which may be streamlined. We can also observe if streamlined genomes are more common among abundant bacteria in the lake, and discover features that are different between streamlined and non-streamlined genomes from the same lake. Analysis of genome streamlining will give a better understanding how important reduction is to be successful in aquatic environments. This work will give a better understanding of genomic streamlining as a microbial lifestyle and how it functions in a community setting.

**To learn about genomic streamlining and its prevalence among abundant lake bacteria, we will characterize the features of genome streamlining in uncultivated bacterial genomes.**

*H3.1: Genome size correlates negatively with growth rate.*

*H3.2: As genome size decreases so do the genes allowing for a diversity of carbon substrate utilization.*

*H3.3: As genome size decreases so do the genes associated with motility or signal transduction.*

*H3.4: Streamlined genomes are traits of highly successful, abundant freshwater microbes.*