Fourth Year Committee Report

Sarah Stevens, McMahon Lab
Meeting: October 26th, 2017 in MSB 5503

I am interested in understanding how bacterial populations diversify and form species. My work aims to better understand the forces that shape bacterial population structure using the lakes as a model system. The McMahon Lab metagenomic time series of many lakes is an ideal dataset for investigating how wild bacterial populations change through time.

Research Progress

Abstract for Accepted Manuscript (ISMEJ)

To understand the forces driving differentiation and diversification in wild bacterial populations, we must be able to delineate and track ecologically relevant units through space and time. Mapping metagenomic sequences to reference genomes derived from the same environment can reveal genetic heterogeneity within populations, and in some cases, be used to identify boundaries between genetically similar, but ecologically distinct, populations. Here we examine population-level heterogeneity within abundant and ubiquitous freshwater bacterial groups such as the acI Actinobacteria and LD12 Alphaproteobacteria (the freshwater sister clade to the marine SAR11) using 33 single cell genomes and a 5-year metagenomic time series. The single cell genomes grouped into 15 monophyletic clusters (termed "tribes") that share at least 97.9% 16S rRNA identity. Distinct populations were identified within most tribes based on the patterns of metagenomic read recruitments to single-cell genomes representing these tribes. Genetically distinct populations within tribes of the acI actinobacterial lineage living in the same lake had different seasonal abundance patterns, suggesting these populations were also ecologically distinct. In contrast, sympatric LD12 populations were less genetically differentiated. This suggests that within one lake, some freshwater lineages harbor genetically discrete (but still closely related) and ecologically distinct populations, while other lineages are composed of less differentiated populations with overlapping niches. Our results point at an interplay of evolutionary and ecological forces acting on these communities that can be observed in real time.
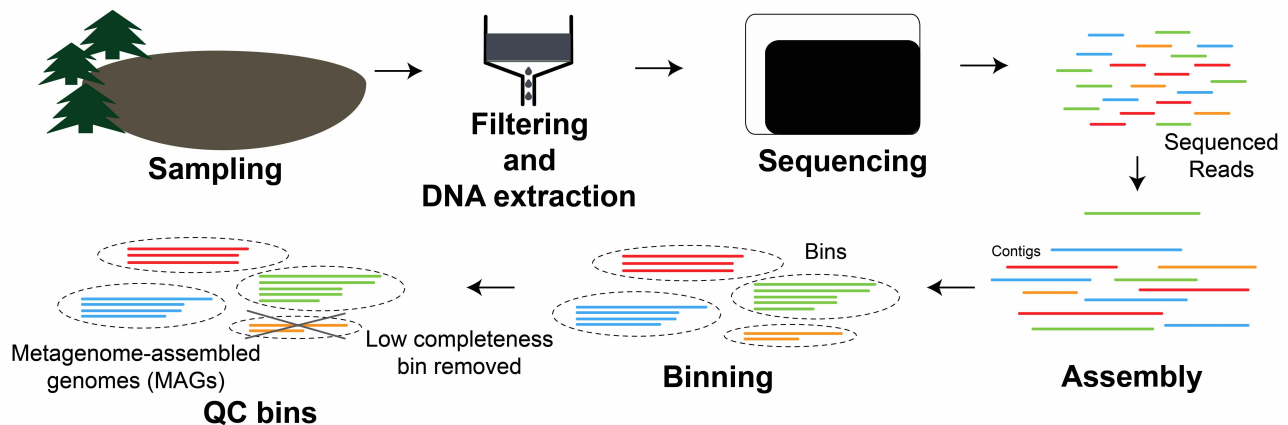
Current Project

Figure 1: Diagram from Sampling to Bins

Questions: Are there related sequence-discrete populations in bog lakes (Crystal Bog, Mary Lake, Trout Bog)? How closely related are they? Do they share a common gene pool? Are there genes present in one lake but not the other for these closely related populations? Are these genepools separated? If so, how did these genes evolve differently between the two lakes?

I have been working on getting bins for Crystal Bog and Mary Lake over the past year. My current challenges are dereplicating them, since we likely have recovered the same genomes from multiple timepoints, and classifying them, as past methods did not scale up. I will also bin additional genomes from Trout Bog using new assemblies of the individual time points. Once I have binned, dereplicated, classified genomes from the three bog lakes I can start comparing genomes within and between lakes. With these genomes, I will search for very closely related genomes and quantify how similar they are across their genomes both the percentage and identity of shared genomic content. For genomes that share at least 96.5% identity across at least 60% of their genome (species level [**Varghese2015**]), I will further investigate what functions are common in shared gene content. If there are few shared species between the lakes this analysis could also be done on genomes from the same genus. This analysis only addresses shared content that assembled in both lakes. I will search for homologs of genes assembled into MAGs in other lakes by mapping the metagenomes from one lake to MAGs from the other. Using this method will both identify homologs present in both lakes and calculate the average similarity of these homologs. I will search for patterns among the shared or absent genes that may be explained by the different environments of the two lakes. I will also look for differences in how the populations evolved by comparing their trees and single nucleotide variants.

New Publications

Garcia, S. L.*, **Stevens, S. L. R.***, Crary, B., Martinez-Garcia, M., Stepanauskas, R., Woyke, T., Tringe, S. G., Andersson, S., Bertilsson, S., Malmstrom, R., McMahon, K. D. (*accepted to ISMEJ*). Contrasting patterns of genome-level diversity across distinct co-occurring populations.

He, S., **Stevens, S. L. R.**, Chan, L.-K., Bertilsson, S., Glavina del Rio, T., Tringe, S. G., . . . McMahon, K. D. (2017). Ecophysiology of Freshwater Verrucomicrobia Inferred from Metagenome-Assembled Genomes. mSphere. doi:10.1128/mSphere.00277-17

*Equal contributors

Conference Poster

Stevens, S. L. R., Egan, R., Malmstrom, R.R., McMahon, K. D. Comparative Genomics of Selected Bacterial Populations from Several Freshwater Bog Lakes. 2017 Microbial Population Biology Gordon Research Conference. July 2017. Andover, NH. Link to Poster

Professional Development

- Software/Data Carpentry
  - Taught 2 Software Carpentry Workshops (1 on-campus, 1 off-campus)
  - Taught 2 Data Carpentry Workshops on-campus
- Attended Anvio workshop at UChicago - April 2017
- Computational Biology, Ecology, and Evolution(ComBEE) group - Started Nov. 2014
  - Taught git workshop (Spring 2017) and gh-pages workshop (Fall 2017)
  - Taught Anvio workshop on campus - May 2017
  - 'Retired' from organizing events - Summer 2017
- Attended Open Science Grid User school - Summer 2017

Awards

- UW-Madison Dept. of Bacteriology Travel Award - Spring 2017