

## Current Project

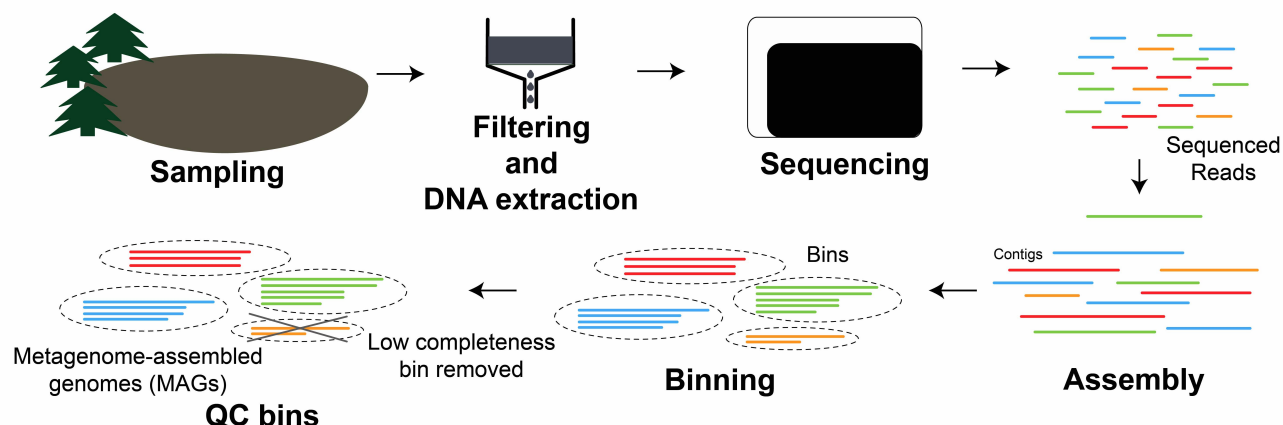


Figure 1: Diagram from Sampling to Bins

I have been working on binning genomes from Crystal Bog and Mary Lake over the past year. One of my current challenges is de-replicating these bins, since we likely have recovered the same genomes from multiple timepoints. Additionally I am working on methods to classify 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. The previously binned genomes from TB were done on a coassembly of all of the timepoints.

Binning and assembling each sample separately provides several advantages which I hope will help to recover more MAG reference genomes. One such advantage is the lower strain heterogeneity in individual samples which can interfere with assembly. Individual assemblies can also better recover genomes from bacteria that were only abundant in one sample but might have been relatively rare in the coassembly which can make assembly difficult. Additionally I hope to be able to recover more of the flexible genome by assembling the same organism in different timepoints.

*Questions: Are there ecologically distinct strains/genotypes within previously defined sequence-discrete populations? What stage of speciation are these distinct strains/genotypes at in their separation? Is there a barrier to recombination between such strains?*

Shapiro & Polz (2014) proposed 5 stages of microbial speciation under differing recombination/selection regimes. In past work we looked for evidence of the two major models of bacterial speciation using the MAGs binned from Trout Bog (Bendall et al., 2016). We used mapped reads from the metagenomic timeseries back to these MAGs and identified operational taxonomic units (OTUs) by delineating sequence-discrete populations at the percent identity threshold where the coverage dropped off. While this method revealed a genome-wide sweep and evidence for past gene-specific sweeps, we could not separate the strains within the sequence-discrete populations since we could not link single nucleotide variants (SNVs) together. New statistical methods have been developed (Costea et al., n.d.; Nayfach, Rodriguez-Mueller, Garud, & Pollard, 2016; Quince et al., 2017) that use models to find linked SNVs and identify haplotypes. Using these methods we can identify strains within the sequence discrete populations and look for ecological differentiation and perhaps catching a variety of stages of speciation put forth by (Shapiro & Polz, 2014). In capturing many stages of speciation we hope to better understand the principles of diversification and how sequence discrete populations emerge and separate from one another.

Since we have multiple metagenomic timeseries from isolate lakes this also allows for some study of allopatric speciation.

## References

Bendall, M. L., Stevens, S. L. R., Chan, L.-K., Malfatti, S., Schwientek, P., Tremblay, J., ... Malmstrom, R. R. (2016). Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISMEJ*.

<https://doi.org/10.1038/ismej.2015.241>

Costea, P. I., Munch 1, R., Coelho, L. P., Paoli, L., Sunagawa, S., & Bork, P. (n.d.). metaSNV: A tool for metagenomic strain level analysis. <https://doi.org/10.1371/journal.pone.0182392>

Nayfach, S., Rodriguez-Mueller, B., Garud, N., & Pollard, K. S. (2016). An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Research*, 26(11), 1612–1625. <https://doi.org/10.1101/gr.201863.115>

Quince, C., Delmont, T. O., Raguideau, S., Alneberg, J., Darling, A. E., Collins, G., & Eren, A. M. (2017). DESMAN: a new tool for de novo extraction of strains from metagenomes. *Genome Biology*, 18(1), 181. <https://doi.org/10.1186/s13059-017-1309-9>

Shapiro, B. J., & Polz, M. F. (2014). Ordering microbial diversity into ecologically and genetically cohesive units. *Trends in Microbiology*, 22(5), 235–247. <https://doi.org/10.1016/j.tim.2014.02.006>