

Investigating Population Dynamics and Genomic Functional Ability of Lake Microbial Communities

Background and Significance:

Many freshwater bacterial lineages, such as the Actinobacteria *acl* lineage, are dominant in the ecosystem but have not historically been studied as they are thus far uncultured (Debroas et al., 2009; Garcia et al., 2013). Most work on bacterial freshwater lineages is based on 16S rRNA gene PCR amplification and sequencing (Newton, Jones, Eiler, McMahon, & Bertilsson, 2011). However, it is becoming increasingly common to sequence the entire DNA collected from an environment, known as metagenomic data or metagenomes (Debroas et al., 2009). Metagenomes have their own limitations, most importantly that in order to track back to specific lineages known reference genomes are needed for comparison.

Previous research used cell sorting, DNA amplification, and sequencing to assemble genomes from single cells (Garcia et al., 2013). Using these Single Amplified Genomes (SAGs) the McMahon lab has been able to look at the abundant freshwater lineage from Actinobacteria called *acl*. This method provides reference genomes, which complement the metagenomic data.

Recently the discovery of genomic streamlining by marine organisms such as the SAR11 clade (Giovannoni et al., 2005; Swan et al., n.d.), has led to the black queen hypothesis that it is advantageous for microbes in the same environment to reduce their genome size by gene loss (Morris, Lenski, Zinser, & Loss, 2012). This is due to some important cellular functions being “leaky”, giving organisms without these genes access to their products (Morris et al., 2012). This genome streamlining might have occurred in some freshwater microbes as well.

Aim 1: Characterize functional potential for common freshwater lineages using SAGs

The McMahon lab has a set of SAGs that represent a variety of freshwater lineages. These genomes are of interest due mostly to their prominence in freshwater settings, but some were chosen due to their divergence from cultured sequenced genomes. The SAGs are all from freshwater lakes, mostly Lake Mendota. I have already done quality control to remove contigs that may be contamination from free-floating DNA or phage. To remove contamination, I examined GC content, tetramer frequency, and phylogenetic BLAST distribution of the contigs.

I am currently investigating functional potential using tools in IMG such as the KEGG, COG and PFAM databases (Chen et al., 2013). I started by searching for the genes whose products are involved in glycolysis, the pentose phosphate pathway, the TCA cycle, oxidative phosphorylation, and cytochromes. In addition, I will look for the genes involved in alternative sugar utilization, polyamine metabolism, phosphorous metabolism, and nitrogen metabolism.

When looking for these enzymes and pathways it is important to understand how complete these genomes are. Single cell genome amplification leads to random bias amplification which results in missing parts of the genome. I will examine each SAG's completeness by taking all of the published, finished genomes for the families of these SAGs and clustering homologs by BLAST results to find all the single copy conserved genes. I will use the presence and absence of these genes to determine the percent completion of each SAGs. From these completion estimates I will be able to calculate the likelihood any gene is missing and apply this to pathway analysis.

This functional information will result in a set of reference genomes for my later work with metagenomes. I will be able to better define niches for these groups and apply the black

queen hypothesis. Overall, this elucidation of function in freshwater settings will expand our knowledge of the metabolism of freshwater community members.

Aim 2: Compare the population dynamics of freshwater microbes over time and environmental conditions and characterize their functional capabilities.

Trout Bog, a humic bog in northern Wisconsin, has been studied in a time series as part of the North Temperate Lakes Long Term Ecological Research (NTL-LTER) and NTL-Microbial Observatory (NTL-MO) projects. These projects have an ongoing time series of both extracted DNA from samples and the corresponding environmental data. This includes 45 metagenomes from both the epilimnion and hypolimnion of Trout Bog, ranging from 2007 to 2009 and 82 and 87 epilimnion and hypolimnion itags(16S) respectively ranging from 2005 to 2009. This data set will be helpful for correlations between community composition and environmental conditions.

Using a previously characterized technique (Wrighton et al., 2012), I intend to assemble genomes from the metagenomes. I will start by assembling the metagenomic reads, either from all samples, a portion of samples, or a single sample from Trout Bog. I will then cluster the resulting contigs into genomes using two methods to discriminate different organisms. First, I will cluster the contigs by plotting their coverage in two different metagenomic samples when the reads are mapped back to these contigs. Contigs from the same genome should have a similar change in coverage between samples. For a second level of separation, I will also make a self-organizing emergent map of the contigs, for each of the coverage clusters, by tetranucleotide frequency, which has been shown to be a distinguishing feature of genomes (Teeling, Meyerdierks, Bauer, Amann, & Glöckner, 2004). To target specific OTUs found to be important using 16S tag sequencing, I will assemble the samples where lineages of interest were more abundant. Ideally, using only these reads should improve the assembly since a greater portion of the metagenomic reads will be from these lineages.

I can then use the method in my first aim to characterize functional capabilities of these genomes. For both the composite genomes from metagenomes and the SAGs, I will look at population changes in the lake over time. Using our time series of metagenomic samples, I will map reads back onto the genomes and see how abundance changes over time and correlates with environmental parameters. From the functional information I will endeavor to find what genes are advantageous given particular lake conditions.

From the functional information I will also make hypotheses for additional experiments. In these experiments I will add specific nutrients to microcosms and identify how the community changes by either qPCR or CARD FISH. I am currently learning CARD FISH with the goal of adding dicarboxylic acid and different amino acids to lake water with my undergraduate student this summer.

Professional Development:

To develop my mentoring skills, I am taking the Delta program's Mentor Training Seminar course. I hope to apply what I learn this summer as I am mentoring my first undergraduate student. I also hope to develop my teaching abilities in the classroom with a Teaching Assistant position in the fall. This summer I will also focus on building my professional network by attending the Microbial Ecology and Water Engineering Conference.

Works Cited

- Chen, I.-M. a, Markowitz, V. M., Chu, K., Anderson, I., Mavromatis, K., Kyrpides, N. C., & Ivanova, N. N. (2013). Improving microbial genome annotations in an integrated database context. *PloS one*, 8(2), e54859.
- Debroas, D., Humbert, J.-F., Enault, F., Bronner, G., Faubladier, M., & Cornillot, E. (2009). Metagenomic approach studying the taxonomic and functional diversity of the bacterial community in a mesotrophic lake (Lac du Bourget--France). *Environmental microbiology*, 11(9), 2412–24.
- Garcia, S. L., McMahon, K. D., Martinez-Garcia, M., Srivastava, A., Sczyrba, A., Stepanauskas, R., Grossart, H.-P., et al. (2013). Metabolic potential of a single cell belonging to one of the most abundant lineages in freshwater bacterioplankton. *The ISME journal*, 7(1), 137–47.
- Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., Bibbs, L., et al. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science (New York, N.Y.)*, 309(5738), 1242–5.
- Morris, J. J., Lenski, R. E., Zinser, E. R., & Loss, A. G. (2012). Evolution of Dependencies through The Black Queen Hypothesis : Evolution of Dependencies through. *MBio*, 3(2), e00036–12.
- Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., & Bertilsson, S. (2011). *A guide to the natural history of freshwater lake bacteria. Microbiology and molecular biology reviews : MMBR* (Vol. 75, pp. 14–49).
- Swan, A. B. K., Tupper, B., Sczyrba, A., Lauro, F. M., González, J. M., Luo, H., Wright, J. J., et al. (n.d.). Prevalent genome streamlining and latitudinal divergence of surface ocean bacterioplankton, (Icm), 1–76.
- Teeling, H., Meyerdierks, A., Bauer, M., Amann, R., & Glöckner, F. O. (2004). Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environmental microbiology*, 6(9), 938–47.
- Wrighton, K. C., Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., Wilkins, M. J., et al. (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science (New York, N.Y.)*, 337(6102), 1661–5.