Dynamics of Sequence-Discreet Bacterial Populations Inferred Using Metagenomics

Sarah Stevens¹, Matthew Bendall², Dongwan Kang², Jeff Froula², Rob Egan², Leong-Keat Chan³, Susannah Tringe², Katherine McMahon², Rex Malmstrom²

¹University of Wisconsin – Madison, Department of Bacteriology
²Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA, USA
USA

March 2014

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231
DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.
**Dynamics of Sequence-Discrete Bacterial Populations Inferred Using Metagenomes**

Sarah Stevens¹, Matthew Bendall², Dongwan Kang², Jeff Froula¹, Rob Egan¹, Leong-Keat Chan², Susannah Tringe², Katherine McMahon¹, and Rex Malmstrom²

¹University of Wisconsin - Madison, Dept. of Bacteriology; ²Department of Energy Joint Genome Institute, Walnut Creek, CA, USA

### Abstract

From a multi-year metagenomic time series of two dissimilar Wisconsin lakes we have assembled dozens of genomes using a novel approach that bins contigs into distinct genomes based on sequence composition, e.g. k-mer frequencies, and contig coverage patterns at various times points. Next, we investigated how these genomes, which represent sequence-discrete bacterial populations, evolved over time and used the time series to discover the population dynamics. For example, we explored changes in single nucleotide polymorphism (SNP) frequencies as well as patterns of gene gain and loss in multiple populations. Interestingly, SNP diversity was purged over time and used the time series to discover the population which represent sequence-discrete bacterial populations, evolved various times points. Next, we investigated how these genomes composition, e.g. k-mer frequencies, and contig coverage patterns at various times points were binned into a single pan-assembly for each environment. Assembled contigs were grouped into genome bins based on sequence composition and coverage patterns we recovered 202 genomes from Trout Bog and from Lake Mendota. We estimate these genomes to be 50-100% complete based on a set of 139 single copy genes conserved among nearly all bacteria. These genomes were then classified based parsing the results of Phylosift(Darling et al., 2014). The results of the classification are summarized in Figure 3.

### Binning Methods

Using the combined assembly of all metagenomes for the hypolimnion of Trout Bog, genomes were initially binned manually by first using sequence-assembly-based classifiers Phylophytis and Classifier for Metagenomic Sequences (ClamO) (Patil, Roune, & McHardy, 2012). Contigs grouped at the family level were further separated into genome bins based on differences in overall coverage. Contigs in the same bin showed a strikingly similar coverage pattern at all time points, thus validating the bins (Figure 2).

### Results and Future Directions

Data for the three previously binned, sequence discrete populations shows all SNPs tending toward fixation in the three-year time series. This provides the first direct evidence supporting the 'ecotype model' of bacterial diversification which predicts selective pressure will periodically purge diversity, genome wide, for an ecotype(Cohan & Perry, 2007). Genes that were gained or lost follow the same pattern as the SNPs, suggesting these genes were present or absent, respectively, in the strain most dominate at the last time points. Next, we intend to investigate the population dynamics of the larger set of genomes binned with Metabat. We will look at SNP diversity across time as well as persistence vs. transience and seasonal dynamics.

### References


