

215. Metagenomic Sequencing of Parallel Consortia Enables Genome Reconstruction of Rare and Closely Related Organisms

William C. Nelson¹* (william.nelson@pnnl.gov), Margie F. Romine¹, Eric Huang¹, Mary S. Lipton¹, Stephen R. Lindemann¹, and James K. Fredrickson¹

¹Pacific Northwest National Laboratory, Richland, Washington

<http://www.pnnl.gov/biology/programs/fsfa/>

Project Goals: The PNNL FSFA is focused on discovering fundamental principles that determine microbial community structure, function, and dynamics, and applying these principles to rational design of microbial communities for desirable outcomes. Our strategy integrates field studies of microbial mat communities and laboratory studies of field-derived consortia and isolates to gain a mechanistic understanding of the environmental and biotic drivers of community higher-order properties such as resistance and resilience. Genome sequence-enabled approaches, including chemical probes directed against specific functions, are being used to elucidate underlying interaction mechanisms. These data are in turn being used to inform community metabolic, spatial and regulatory network models. Our research plan supports DOE goals to achieve a predictive understanding of microbes and microbial communities and to provide foundational knowledge enabling rational design of microbial systems.

Metagenomic sequencing has enabled the de novo reconstruction of microbial genomes from environmental communities. Biological complexity has limited this approach to low-diversity communities and abundant members of complex communities. The genomic complement of each member population must be determined, however, to gain a predictive understanding of the dynamics of interactions between populations and the contributions of these interactions to overall community function.

We performed short-read metagenomic sequencing of two unicyanobacterial consortia cultivated from saline (MgSO₄) phototrophic microbial mats (previously described in [1]) and reconstructed member species' genomes. These are the most complex communities for which near-complete, species-resolved (and in some cases strain-resolved) genomic information has been generated. Near-complete genome sequences were obtained for 17 of the 20 member species found in the two cultures. Our use of parallel consortia (consortia cultivated under the same conditions from the same natural community and having overlapping species composition) allowed acquisition of near-complete genome sequence for organisms that are rare in the natural community, and detection and deconvolution of microdiversity. Differential membership and abundances between the two consortia revealed inter- and intraspecies microdiversity, resolving two *Halomonas* sp. and two *Rhodobacteraceae* sp. ecotypes previously clustered by amplicon analysis. Differences in functional potential (~850 genes in the *Halomonas* spp.) between the resolved organisms suggest distinct roles in the community. This work demonstrates how significant functional diversity can be hidden within an OTU, and how subtle variations between genomes might result in distinct niches.

In addition, we have been able to assess the accuracy of common practices in genome reconstruction, such as scaffold segregation by differential read coverage and conserved single-copy gene analysis, through comparison of our reconstructed genomes against available complete genome sequences from isolates cultured in parallel to the metagenomic study. Comparison of the reconstructed genomes to the complete genomes of isolated strains demonstrated the sensitivity and specificity of the reconstruction process across a wide range of sequence coverage (9x – 2700x).

The utility of such datasets in interpreting environmental data was demonstrated by screening an environmental proteomic data set against the reconstructed genomes and identifying peptides from 19 of the organisms in the consortia.

Our work represents a critical step toward understanding the mechanisms that drive microbial community formation, function, and response to environmental stimuli. The genomic information generated allows prediction of the functions of member taxa (see Posters by M. Romine and S. Lindemann) and provides a foundation against which experimental transcriptomic and proteomic data can be evaluated. By enabling in toto community genomic reconstruction, the parallel consortia approach serves as a useful intermediate between environmental metagenomics and single-cell amplified genomics.

References:

1. Cole JK, Hutchison JR, Renslow RS, Kim YM, Chrisler WB, Engelmann HE, Dohnalkova AC, Hu D, Metz TO, Fredrickson JK, Lindemann SR: Phototrophic biofilm assembly in microbial-mat-derived uncyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions. *Front Microbiol* 2014, 5:109.

This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genome Sciences Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL). Sequencing was done at the DOE Joint Genome Institute under Contract No. DE-AC02-05CH11231 and Community Science Project 701.