Composition of Metagenomes from Yellowstone Hot Spring Microbial Mats Constructed by Photosynthetic Prokaryotes

Christian G. Klatt[1], Jason Wood[1*], Doug Rusch[2], Donald A. Bryant[3], William P. Inskeep[1], and David M. Ward[1]

[1]Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT; [2]The J. Craig Venter Institute, Rockville, MD; and [3]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, State College, PA

We have analyzed metagenomic data obtained from microbial mats constructed by cyanobacteria in alkaline siliceous hot springs and by anoxygenic phototrophic bacteria in sulfidic carbonate hot springs of Yellowstone National Park. Metagenomic libraries were analyzed from four samples—the top green layers of ~60-65°C cyanobacterial mats from Octopus and Mushroom Springs, the undermat layers from the 60°C Mushroom Spring mat, and mats from Bath Lake Vista Annex (BLVA) Spring taken at two different seasonal time points. We have used genomes for representative isolates cultivated from these or similar systems in BLAST recruitment analyses. The majority of metagenomic sequences from the top green layers of cyanobacterial mats were recruited by genomes of phototrophic isolates (*Synechococcus* sp. A and B', *Roseiflexus* sp. RS1, *Chloroflexus aurantiacus* sp. J-10-fl, Y-400-fl and 396-1, *Candidatus* Chloracidobacterium thermophilum and *Chloroherpeton thalassium*), with much lower recruitment by genomes of nonphototrophic isolates. Examination of % nt identity of recruited sequences revealed that these mat layers contain populations that are: (i) very closely related (i.e., >90% nt identity) to *Synechococcus*, *Roseiflexus* and *C. aurantiacus* 396-1, (ii) more distantly related (70-90% nt identity) to *Candidatus* C. thermophilum, *C. thalassium*, and other *Chloroflexus* strains, (iii) so distantly related to reference genomes that recruitment may have been fortuitous and (iv) not related to any of the reference genomes. By aligning paired sequences recruited by the *Synechococcus* sp. A genome in the BLAST analysis against this genome as a function of % nt identity, we noted the presence of sequence subpopulations that are either extremely closely related or more distantly related to this reference genome. Synteny increased with relatedness of metagenomic sequences to genomic homologs. Sequences related to *recA* exhibited a similar view of metagenome composition with high representation of dominant phototrophs, but also indicated the presence of *Aquificales*, *Firmicutes*, and multiple divisions within the Proteobacteria. An independent oligonucleotide composition analysis was done with scaffolds constructed with metagenomic sequences by the Celera assembler. Isolate genomes and scaffolds greater than 5kb in length were profiled using the frequency distribution of all possible tri-, tetra-, penta-, and hexa-nucleotides and these profiles were compared using a principal components analysis. Scaffolds with similar oligonucleotide composition showed evidence of similar taxonomic origin, as scaffolds containing phylogenetic marker genes grouped with genomes of similar oligonucleotide character. Most scaffolds were grouped using a k-means clustering algorithm into clusters containing the isolate genomes used as references. Scaffolds grouping in clusters not containing reference genomes may represent uncharacterized community members that could be targeted for future analysis. Additional metagenomic libraries from Mushroom Spring bottom layers were subjected to a similar BLAST recruitment analysis. These BLAST results indicated that the lower layers of this mat community are dominated by the presence of organisms closely related to *Roseiflexus* spp., with decreased representation of *Synechoccocus* spp. Metagenomic libraries originating from BLVA had very little representation from oxygenic phototrophs, and were dominated by sequences from *Chloroflexus* and *Roseiflexus* spp. Future work will focus on the differences in diversity patterns of anoxygenic phototrophs among these metagenomic libraries.