

Poster #155

Genome-Resolved Metagenomic and Geochemical Analysis of East River Riparian Zone Soils Supports the 'Systems Within Systems' Approach for Watershed Analysis

Paula B. Matheus Carnevali¹, Kenneth H. Williams², Wenming Dong², Brian C. Thomas¹, Susan Hubbard² and Jillian F. Banfield^{1,2,3}

¹ Department of Earth and Planetary Sciences, University of California, Berkeley

² Lawrence Berkeley National Laboratory

³ Department of Environmental Science, Policy and Management, University of California, Berkeley

Contact: Jill Banfield [jbanfield@berkeley.edu]

Microorganisms largely control carbon turnover and mediate other important biogeochemical processes in soils. Information regarding microbial impacts at larger spatial scales is required to understand ecosystem processes. Our approach to scaling genome-resolved metagenomics to study the riparian zone was to target discrete reaches of the East River as a potentially representative repeating motif. We are testing this hypothesis using extensive sampling and DNA sequencing based analyses. 94 soil samples were collected over the 10-25 cm depth interval from floodplains associated with three meandering reaches: upstream (MG), midstream (ML), and downstream (MZ). Genomic DNA was extracted and then sequenced at the Joint Genome Institute on a HiSeq Illumina instrument (~ 5 Gbp per sample).

DNA sequence information from individual samples was processed and assembled using standard methods. Based on profiling of community composition using scaffolds encoding *rpS3*, we identified 813 distinct strains, many of which were found at multiple sites within a meander and also occurred in two or all three meanders. Shortly after assembly we have already recovered 164 draft or higher quality genomes representing 88 distinct microorganisms. Most of the relatively abundant organisms in the three systems are represented by draft genomes, the main exception being some members of a group of highly abundant and closely related strains of Betaproteobacteria. For many organisms, the same genome was recovered from multiple samples. For example, for one novel Betaproteobacteria we reconstructed 11 near complete genomes and a genotype of Gammaproteobacteria is represented by 11 genomes. Other groups of abundant, novel, and genomically defined Bacteria include Acidobacteria, Ignavibacteria, Nitrospirae, Gemmatimonadetes, Deltaproteobacteria, Verrucomicrobia, and Planctomycetes. Thaumarchaeota (Nitrosopumilales) were the most abundant Archaea observed in these samples. Additionally, we recovered genomes for a variety of very little understood organisms known only from prior genomic studies, mostly at the Rifle SFA research site. These include Rokubacteria, WS3X, CHLX, Zixibacteria, OD1 (various Parubacteria), and DPANN (Pacearchaeota). Widely distributed functions in addition to carbon turnover include sulfur oxidation, iron oxidation, ammonia oxidation, nitrate reduction, and nitrogen fixation. Despite heterogeneity in species distribution, strong overlaps in organism membership and similar functions are well represented across the three localities; however, the function of some organisms may vary across sites. Overall, we conclude that genome-resolved metagenomics, analyzed in the context of physical and chemical information, can be scaled to understand the distribution of ecosystem processes. Insights regarding community composition from a single floodplain may be generalizable, at least under some conditions.