DOE Hanford Site Metagenome: A multiple extreme environment that hosts wide diversity of microbes and radiotolerant bacteria

Presented is the analysis of the metagenome of the living microbial community extracted from low biomass (~10,000 cells per gram) subsurface sediments (60 to 150 feet deep) beneath a leaking high-level radioactive waste tank at the DOE Hanford Site. Low-coverage shotgun sequencing was performed by the DOE Production Genomics Facility. The results of these analysis are available at http://compbio.mcs.anl.gov/PNNL1

63 16sRNA and 13,388 protein sequences from four contamination zones were analyzed.

Higher contamination zone was dominated by *Deinococcus-Thermus*, *Actinobacteria* and *Firmicutes*. 60% of protein bi-directional best Blast hits were against the sequences from Deinococcus-Thermus phylum and 40% of hits from Actinobacteria. 25 of 40 16sRNA were attributed to Actinobacteria and Firmicutes. Micrococcineae, Propionibacterineae, Corynebacterineae and Frankineae suborders within the Actinobacteria were represented in the higher contamination zone samples. The samples also contained a significant number of homologs to proteins from extremophilic organisms, mostly originating from Firmicutes and Archaea.

Low contamination zone was dominated by Proteobacteria and Actinobacteria. 56% of protein bi-directional best Blast hits were against the sequences from Proteobacteria phylum and 40% of hits from Actinobacteria. 6 of 9 16sRNA were attributed to Actinobacteria species, specifically Micrococcineae, Propionibacterineae and Corynebacterineae suborders.

Relatively low percentages of amino acid identities (and e-values) in higher contamination zone for hits to the Actinobacteria and Firmicutes and some orders of Proteobacteria in the low contamination zone suggests presence of a novel organisms or even orders of Actinobacteria and Proteobacteria in this metagenome in comparison to currently available sequenced species.

Metabolic Reconstructions from sequence data.

Metagenome proteome contained genes potentially representing 178 distinct enzymatic function in lower contamination zone and 186 distinct enzymatic functions in higher contamination zone. These enzymatic functions were "binned" to 62 organisms in higher contamination zone and 56 organism in lower contamination zone. However due to poor coverage and sparse character of the metagenomic data, we have merged data corresponding to particular organisms into the larger "genus" bins combining all the sequences originating from a particular genus.

The presentation will provide more details regarding the methods and approaches used for analysis of Hanford metagenome and provide illustrative examples of the results.