Poster M-16

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

Authors:
Syed H Mustafa (*Argonne National Laboratory*)
Natalia Maltsev (*Argonne National Laboratory*)
Fred Brockman (*Pacific Northwest National Laboratory*)
Banu Gopalan (*Pacific Northwest National Laboratory*)

Short Abstract: Metagenome analysis of the microbial community residing in low biomass subsurface sediments beneath a leaking high-level radioactive waste tank at the DOE Hanford Site revealed that it was dominated by representatives of Deinococcus, Actinobacteria and Firmicutes in higher contamination zones, while Proteobacteria and Actinobacteria dominated in lower contamination zones.

Long Abstract:

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

The following analyses of metagenome were performed

1. Identification of the taxonomic origin of the species in metagenome using 16S RNA. Sequence analysis of the 16S rRNA gene represents a highly accurate and versatile method for bacterial classification and identification, even when the species in question is notoriously difficult to identify by phenotypic means. The 16s rRNA sequences extracted from the Hanford site were analyzed using Blast against RDP database to estimate taxonomic positions of the resident organisms. .
2. Analysis of Metagenomic proteome to identity potential functions of the genes represented in the samples. The metagenomic sequence data was analyzed by a number of conventional bioinformatics tools (e.g. BlastX, Blocks) and annotated with the GO terms. Analyses of large volumes of metagenomic data by a variety of bioinformatics tools requires substantial computational resources. We have used the GADU workflow pipeline being developed by our group and Grid resources of TeraGrid and Open Science Grid to perform these analyses.

3. Identification of taxonomy-specific variation of enzymes using the Chisel workbench (http://compbio.mcs.anl.gov/CHISEL) being developed by the Bioinformatics group at ANL. Such analysis allows developing or refining the suggestions regarding the taxonomy of the organisms found in the metagenome and attribute particular enzymatic steps to specific taxonomic groups. Identification of taxonomic variations of enzymes allows the development of function and taxonomic-group specific oligonucleotides to be used in wet lab analysis. Experimental identification of “signature” enzymes or groups of enzymes in metagenome will

assist in identification of metabolic processes characteristic for the sequenced microbial community.
4. Metabolic reconstructions. This project leverages the PUMA2 knowledge base and tools for the development of the metabolic reconstructions from sequence data. The gene functions, attributed to particular organisms or taxonomic groups of organisms were projected onto the library of metabolic pathways from the EMP and KEGG databases. Metabolic models were constructed for the organisms or taxonomic groups found in the samples. They represent valuable scientific hypotheses regarding an organism's physiology and facilitate experimental planning.

Results

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 results At least one gene was present for synthesis of 18 of the 20 amino acids. Pathways in which ten or more genes in the pathway were present include metabolism of purines, pyrimidines, aminoacyl-tRNA, glycolysis/gluconeogenesis, pyruvate, starch, glycerolipid, porphyrin, glycine/serine/threonine, arginine/proline; and phenylalanine/tyrosine/tryptophan.

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