The use of ‘Metaomic’ and Modeling Approaches at Subsurface Sites to Study Carbon, Nutrient and Metal Cycling in Microbial Communities

Kelvin Li1, Monika Bihan1, Amy Boaro2, Neha Gupta1, Christopher Gowen3, Jiao Zhao3, Kenneth H. Williams4, Mary Lipton2, Krishna Mahadevan3, Barbara A. Methé1* (bmethe@jcvi.org)

1J. Craig Venter Institute, Rockville MD, 2Pacific Northwest National Laboratory, Richland WA, 3University of Toronto, Canada, 4Lawrence Berkeley National Laboratory, Berkeley CA

**Project Goals:** Diverse microbial communities exist in subsurface environments that possess significant metabolic potential to effect global carbon, nitrogen and metal cycles including the transformation of radionuclides. Objectives of this ongoing project are: 1) to apply systems-level biology through application of ‘metaomics’ approaches (collective analyses of whole microbial community DNA, RNA and protein) to the study of microbial environmental processes and their relationship to C, N and metals including the influence of microbial communities on uranium contaminant mobility in subsurface settings undergoing natural attenuation, 2) improve methodologies for data generation using metaomics technologies and analysis and interpretation of that data and 3) use the data generated from these studies towards microbial community-scale metabolic modeling (see also “Genome-scale Metabolic Modeling of Microbial Communities at Sites Undergoing Natural and Stimulated Attenuation of Uranium”).

To meet the goals of this ongoing project, two subsurface sites from the Department of Energy (DOE) Rifle Integrated Field Research Challenge Site (RIFRC) are being interrogated using a suite of metaomic approaches. The first site consists of sediments from the Winchester 2007 gallery, ‘JB” well locations and was chosen due to the occurrence of natural attenuation of uranium (uranium reduction in the absence of biostimulation or other remedial interventions). The second and more recent sites of study within this project have been collected from Colorado River Floodplain (CORFP) sediments representing recent sediment depositions. Overbank deposits in the floodplain have become enriched in C, Fe and S minerals. Aggradation processes have led subsequent burial of these enriched sediments creating “hotspots” of biogeochemical activity which serve as analogs to the buried naturally reduced sediments at the JB sites.

Previous applications of metagenomic (DNA) and metatranscriptomic (RNA) investigations from the ‘JB’ sediments have been generated using the Illumina HiSeq and MiSeq platforms Taxonomic profiles generated from both assemblies and high quality alignments (>60bp quality trimmed read length at >80% composite identity) to the NCBI NT database revealed that for both the metagenomic and metatranscriptomic data sets, the most abundant species based on best matches (~28% DNA, ~14% RNA) are to relatives of the facultative anaerobic chemolithotroph, Thiobacillus denitrificans, capable of coupling the oxidation of inorganic sulfur compounds to the reduction of oxidized nitrogen compounds and oxidation and potential remobilization of U(IV) to soluble U(VI) in the presence of nitrate. Evidence for the presence of metal reducing bacterial relatives (although not necessarily demonstrated to reduce uranium) such as Rhodoferax ferrireducens, were also determined.

The relatively high abundances of T. denitrificans and R. ferrireducens in the JB data served as motivation to develop a single combined genome-scale model based on a constraint-based metabolic reconstruction of the two organisms. An automatically generated model (through Model SEED) was available for T. denitrificans which required additional manual curation to facilitate its use in this study. Both individual models have been evaluated for their ability to produce flux through their respective biomass reactions after which the stoichiometry matrices underlying each organism's model were then combined. Among the findings from this model are that as the proportion of the biomass objective function shifts in dominance from 100% R. ferrireducens to 100% T. denitrificans, the flux through the combined biomass reaction increases towards 50%, suggesting a possible synergistic instead of
competitive relationship between the two organisms. Other outcomes include the finding of an increase of maximum flux towards T. denitrificans which may suggest a slightly more flexible set of capabilities for obtaining carbon and energy from a minimal soil media.

The CORFP samples represent an important opportunity to contrast microbial community diversity and function especially the coupling of carbon, nutrient and metal cycles with the results obtained from the JB sites using metatranscriptomic and metaproteomic approaches. To that end, biological and technical replicates of the metatranscriptome from both Illumina (short-read, ~100bp) and PacBio (long-read, ~3kb) sequencing platforms along with metaproteomic data from LC MS-MS approaches are currently being analyzed. Both data types support the finding that a significant proportion of the microbial community has been assigned to the beta- proteobacteria, including T. denitrificans, similar to the taxonomic profiles at the JB sites.

Investigations of important biological functions are underway using both data types which include support for the activity of processes related to energy metabolism via glycolytic pathways and oxidative phosphorylation. For example, highly expressed genes from metatranscriptomic data were determined by using a statistical analysis in which the abundances of a collection of housekeeping genes was used to compute a null distribution to which the relative abundance of transcripts were compared to identify those genes whose expression was significantly greater (p-value <0.05) than the housekeeping genes. Significant expression was determined for genes such as NADH dehydrogenase (oxidative phosphorylation) and glyceraldehyde phosphate (glycolysis). Using metaproteomic approaches, 620 proteins (FDR=8) were identified from a total of 3144 spectra. The most abundant proteins include ATP synthase subunits (n=58), and K(+)‐insensitive proton pumps (n=11) (oxidative phosphorylation) as well as enolase (n=12) (glycolysis), and glyceraldehyde‐3‐phosphate dehydrogenase (n=11). The metatranscriptomic data also revealed significant expression of genes related to functions including heavy metal efflux pumps and other transporters, and PAS domain containing genes important to sensing environmental conditions. While metaproteomic data supported the presence of an active sulfur cycle including proteins from T. denitrificans, related to sulfur oxidation as well as proteins associated with sulfate reducers from the family Desulfobacteraceae (n=32). Currently, qRT-PCR approaches are also underway to validate metatranscriptomic results. Collectively, these investigations provide an opportunity to identify both the specific and potentially unique contributions of each metaomic data type (metascriptome and metaproteome), including their individual sensitivities and specificities along with the nature and extent of the added value from data integration.

This project is supported by the Office of Biological and Environmental Research in the DOE Office of Science DESC0006837.