

29. Microbial Community Structure Predicts Groundwater Geochemistry

Andrea M. Rocha^{1,2*} (rochaam@ornl.gov), Mark Smith³, Christopher Smillie³, Julian L. Fortney^{1,2}, Stephen M. Techtman^{1,2}, Dominique Joyner^{1,2}, Tonia L. Mehlhorn¹, Jennifer E. Earles¹, Kenneth A. Lowe¹, Janna Phillips¹, David B. Watson¹, James H. Campbell¹, Susan Pfiffner^{1,2}, Kati Ayers², Charlie Paradis², Joy D. Van Nostrand⁴, Liyou Wu⁴, Ping Zhang⁴, Zhili He⁴, Jizhong Zhou⁴, Michael W. W. Adams⁵, Andrew Lancaster⁵, **Paul D. Adams**⁶, Adam P. Arkin⁶, **Eric J. Alm**³ and **Terry C. Hazen**^{1,2,6}

¹ Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, ² University of Tennessee, Knoxville, TN, ³ Massachusetts Institute of Technology, Cambridge, MA, ⁴ University of Oklahoma, Norman, OK, ⁵ University of Georgia, ⁶ Lawrence Berkeley National Laboratory, Berkeley, CA,

<http://enigma.lbl.gov>

Project goals: ENIGMA's Field Microbiology component works with microbial communities in the environment to investigate how biological structure and function relate to critical biochemical conditions/changes, such as the immobilization of toxic metals and the flow of carbon and nitrogen. The goal is to identify key microbial populations and determine the community events and mechanisms of these populations that impact and control environmental activities of interest, ultimately predicting how perturbations of the environment may affect community structure. From these results, we can develop models that can be applied to microbial populations overlaid with geochemical parameters and engineering controls. Development of such a model and key findings from a 100-well survey at the Department of Energy's Oak Ridge Field site is described below. Specifically, we demonstrate the ability to utilize microbial community assembly from independent groundwater environments to accurately predict the geochemistry and elucidate key systems biology features.

One of the primary objectives of the ENIGMA Environmental Core Field Microbiology Component is to design an efficient field sampling study that will maximize the geochemical diversity of the study site and provide for enhanced resolution of microbial communities and geochemical associations. At the Department of Energy's Oak Ridge field research site, 243--acres of contaminated area is located within the Y--12 plant area of responsibility of the Oak Ridge Reservation (ORR). Here, over 20 years of historical and published data for more than 800 groundwater wells is available in a computer queryable database. In this study, we conducted a survey of 100 groundwater wells in order to (1) characterize key microbial populations at geochemically distinct locations and (2) identify associations between environmental gradients and microbial communities. We hypothesize that differences in geochemistry underlie differences in microbial community structure across the groundwater wells. In order to identify microbe-environment associations, a statistically informed experimental design was developed by ENIGMA's computational core. To optimize geochemical diversity and identify wells where environmental factors are uncorrelated, k-medians clustering algorithm was used to group 818 wells into 100 clusters using 14 geochemically similar measurements. At each well, *in situ* groundwater measurements were recorded and unfiltered and filtered groundwater samples were collected for both geochemical measurements and analysis of microbial communities. Nucleic acids were collected by filtering 4-L of water through a 10.0µm pre-filter and 0.2µm-membrane filter and then extracted using a Modified Miller method. Community genomic DNA yields from the 0.2µm and the 10.0µm filters range from 0.096- 8.5µg and 0.096-22µg, respectively. Cell counts for the field samples ranged from 10³-10⁶ cells/mL and were consistent with biomass estimates from phospholipid fatty acid analysis (PLFA). Initial analysis of PLFA data suggests stress indicators for Gram Negative populations, such as those associated with pollutants and nutrient limitation, are present in 15% of the wells sampled. Evaluation of

divergence of microbial communities across all the wells indicates the microbial communities are fairly distinct. Comparison of microbial communities within each well shows taxa are not as divergent compared to across all wells. However, initial analysis indicates there are some organisms unique to the 10.0 μ m size fraction. A total of 27 wells, which can be classified into 7 groups based on pH and contaminant concentrations, were analyzed with the newest GeoChip 5.0_180K to evaluate functional differences and potentials of microbial communities under different environmental conditions. Our results indicated that the groundwater microbial community functional structure was significantly different among seven different groups of wells, and such community differences were largely correlated with the differences in pH, U(VI), nitrate, dissolved organic carbon (DOC) and sulfate in the groundwater. The microbial functional diversity significantly decreased with U concentrations. Metadata correlations of all the wells show many of the geochemical parameters are independent of each other. Using the synthetic learning in microbial ecology (SLiME) algorithm and the large independent dataset, we are able to predict the geochemistry from the 16S rRNA. Additionally, using the relationship between the geochemistry and microbiology, the critical OTUs that geochemistry can be identified. Overall, results from this study indicate that with careful design and a large dataset, the groundwater microbial community structure can be used to accurately predict the geochemistry. Such predictions may provide the ability of microbial monitoring for natural attenuation at legacy sites and be enabling for ENIGMA for more specialized questions on microbial community and network structure and function.

This work conducted by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (<http://enigma.lbl.gov>), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory, was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02- 05CH11231.