Integration of metagenomics and consortia data to study microbial interactions and community assembly

L.M. Lui[1*] (lmlui@lbl.gov), H.K. Carlson[1], A.W. Sczesnak[1,2], O. Erbilgin[1], M. De Raad[1], V. Mutalik[1], A. Kazakov[1], X. Wu[1], A. Aaring[1], J. Voriskova[1], J. Kuehl[1], M. Price[1], D. Chivian[1], N.B. Justice[1,2], T. Simmons[1,2], K. De Leon[3], Dominique Joyner[4,5], H. Smith[6], P.S. Novichov[1], A.M. Deutschbauer[1], R. Chakraborty[1], T.R. Northen[1], J.D. Wall[3], T.C. Hazen[4,5], M.W. Fields[6], **A.P. Arkin**[1,2] and **P.D. Adams**[1,2]

[1]**Lawrence Berkeley National Lab, Berkeley**; [2]University of California at Berkeley; [3]University of Missouri, Columbia; [4]University of Tennessee, Knoxville; [5]Oak Ridge National Lab, Oak Ridge, [6]Montana State University, Bozeman;

http://enigma.lbl.gov

Project Goals: ENIGMA (Ecosystems and Networks Integrated with Genes and Molecular Assemblies) uses a systems biology approach to understand the interaction between microbial communities and the ecosystems that they inhabit. To link genetic, ecological, and environmental factors to the structure and function of microbial communities, ENIGMA integrates and develops laboratory, field, and computational methods.

ENIGMA uses a systems biology approach to build a predictive understanding how phenomena at the genetic, community, ecological, and environmental level influence microbial community assembly and activity. Sampling of sediment and groundwater microbial communities from the Oakridge National Laboratory Field Research Center (ORNL FRC) are coordinated with measurements of biogeochemical parameters to provide a rich set of samples, isolates, and data for integrated analysis across ENIGMA labs. Three categories characterize our efforts to mechanistically understand microbial community assembly: (1) direct analysis of isolate interactions and use of genetic tools to study them, (2) enrichment of communities under controlled laboratory conditions, and (3) analysis of amplicon and whole genome shotgun sequencing data. Each of these types of analyses is informed by the biogeochemical measurements from the field site and functional and genetic fitness characterization of isolates (see other ENIGMA posters).

We study microbial interactions directly with synthetic communities and are developing novel microfluidics and isolation methods to directly observe interacting cells. Work with a consortium of *Pseudomonas* spp., the most common type of isolate from the FRC, has revealed that the type of carbon source available can influence the production of inhibitory secondary metabolites. Analysis of metabolomics data (NIMS and RP-LC-MS) derived from spent media provides candidate inhibitory compounds and the program MAGI is used to predict genes that produce them. Growth of transposon mutant libraries in spent media can provide information about which genes lead to susceptibility to these compounds. Preliminary data from experiments with combined carbon sources suggest that there is a “tipping point” in the proportion of a particular

carbon source that is needed for an isolate to produce an inhibitory effect on other microbes. Currently, we are developing methods to contain sediment particles or transposon mutant libraries in microfluidics droplets to observe microbial interactions and irreducible communities. We are also developing a method that studies flocs as we hypothesize that in ORNL FRC groundwater direct interactions of microbes are likely to occur within flocs of two or more cells. We are collecting flocs for identification of its members as well as for isolations and enrichments that will be used to reconstruct these floc communities in the laboratory. To provide a means to edit microbes in a group and dissect ecosystem function, we also have a discovery project to find bacteriophages at ORNL FRC.

Cultivation of enrichments provides a means to study sample communities under field relevant conditions and to study how ecological and environmental selection pressures in the same experiment affect community assembly. Building from experiments of isolates obtained with specific carbon sources, an experiment with 293 microcosms inoculated with ORNL FRC groundwater with varied type and number of sugars as a carbon source suggests that (1) a higher diversity of carbon sources leads to higher species diversity and (2) some carbon sources have a higher selection effect on the community structure than others. To study sulfate and nitrate reduction, two important metabolic activities at ORNL FRC, we are developing techniques for cultivation of enriched microbial communities under sulfate and nitrate reducing conditions. Enrichments are initiated under anoxic conditions using a variety of field relevant carbon sources. Additionally, we are using field geochemical information to guide media design.

Finally, we are using amplicon and whole genome metagenomics shotgun sequencing to study community structure across ORNL FRC and to study predicted microbial functional profiles in conjunction with field data. Comparison of the 16S rDNA sequences that were most abundant at the field site (either in groundwater or sediment) to the sequences of >1,000 isolates from the site informs our isolation efforts. We predict that that many of the uncultivated yet abundant microorganisms are aerobic chemolithoautotrophs or slow-growing facultative heterotrophs that utilize more recalcitrant carbon. We have also developed a computational pipeline to do a fast, focused analysis of nitrogen cycle related genes in metagenomics data and used it to analyzed six samples that represented sections of a 20 foot sediment core. This pipeline can be adapted to other functional roles. This pipeline yielded a view of how nitrification changes with depth and predicts a consortium of an ammonia-oxidizing thaumarchaeon and a nitrite-oxidizing bacterial species. In addition to this computational pipeline, we are working with KBase to integrate other metagenomics tools for fast analysis of other ENIGMA samples. Currently we are working on metagenomics of two more cores and accompanying groundwater from ORNL FRC that are coordinated with biogeochemical data and activity data.

ENIGMA is a Scientific Focus Area Program at Lawrence Berkeley National Laboratory and is 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.