ENIGMA – Biotechnology: High Throughput Chemogenomic Fitness Profiling of ENIGMA Relevant Microbes Using a Next Generation Barcode Sequencing Assay

Adam M. Deutschbauer[1], Jennifer He[1], Morgan Price[1], Jennifer V. Kuehl[1], Kelly Wetmore[1], Nancy L. Liu[1], Sonia Reveco[1], **Swapnil R. Chhabra**[1], Grant M. Zane[2], Tom R. Juba[2], **Judy D. Wall**[2], **John-Marc Chandonia**[1], **Trent R. Northen**[1], **Adam P. Arkin**[1] and **Gareth P. Butland**[1*] (GPButland@lbl.gov)

[1]Lawrence Berkeley Natl. Lab, Berkeley, CA, [2]Univ. of Missouri, Columbia, MO

Project Goals: ENIGMA's goal is to understand, at a molecular systems level, the bacterial soil communities at DOE sites contaminated with heavy metals or radionuclides. Environmental change or human intervention will alter the chemical environment in the subsoil, which in turn affects which species predominate as well as microbial physiology. It is therefore critical to understand how microbes respond to environmental changes which may invoke a cellular stress response and inhibit growth or alternatively provide an opportunity to thrive within a community. A comprehensive understanding of the pathways impacted by specific stresses or changes in environment, together with data from other ENIGMA groups detailing changes in community composition as a whole under similar conditions will allow us to link the viability of specific organisms of interest to their ability to contribute or perform specific functions for or within the community.

To ascertain the key processes contributed by an organism to a community and to characterize that process within the organism in molecular detail, it is necessary not only to identify the components of the pathways involved, but to also understand their functional interactions with other processes in the organism of interest. To address these questions, we have utilized recently created comprehensive genome wide collections of single mutant knockout strains and performed chemogenomic fitness profiling to assess the change in viability (growth fitness) of thousands of mutant strains under a variety of growth conditions (i.e. presence of stress agent, change in growth medium, etc.) in a high throughput format. Chemogenomic interactions are identified when a specific mutant or collection of mutant strains have a statistically significant defect or growth advantage under a specific condition when compared to other mutants, and are a powerful indicator of functionally related genes which may be present in the same or related pathways and processes.

Previously used methods to assay the fitness of mutant strains in model systems have relied on the imaging of colony growth on agar plates. This method is not an option for a potentially diverse range of microbes of interest to ENIGMA which may not form morphologically uniform colonies or may excrete chromophoric substances into the media, precluding accurate imaging. However, due to the successful development and application of a TagModule based approach to create mutant strains of *Shewanella onedensis* MR-1, *Desulfovibrio alaskensis* G20 and *Desulfovibrio vulgaris* Hildenborough with dual unique 20nt molecular barcodes, we are leveraging these strain collections to perform pooled fitness assays, containing thousands of mutants in a single sample. Microarray hybridization or next generation sequencing can be used as a readout for the change in abundance of specific mutant barcodes within the pool under specific stress conditions.

To this end, we have recently designed and implemented an assay, based on pioneering work in yeast, which utilizes the sequencing of unique molecular barcodes on an Illumina platform as a quantitative readout for mutant strain fitness. This assay, as with a microarray readout, is

species independent but will allow for massively increased throughput and dynamic range compared to existing approaches at a lower per experiment cost. Initial development of this assay utilized a collection of *S. oneidensis* mutants, each containing a different pair of molecular barcodes, due to the large quantity of microarray based chemogenomic fitness data available from this collection for comparative analysis. Preliminary assay data comprised over 173 million usable barcode reads for a single Illumina lane. This massive sequencing capacity has allowed us to utilize experimental indexes to analyze multiple conditions on the same Illumina lane. Using an initial 18-index multiplex, for an ~4000 mutant complexity pool, we were able to obtain an average of ~2000 barcode reads for each mutant present (typical range 700 to 3000). These data and previous work in yeast suggests that at minimum, 48-index multiplexing (24-48 stress conditions per lane) is feasible. As our Illumina library preparation protocol optimization is complete, we will apply this technique to perform chemogenomic fitness profiling of single gene mutants of *Desulfovibrio alaskensis* G20 and *Desulfovibrio vulgaris* Hildenborough to highlight the hypersensitivity and resistance of strains to stress conditions, chemical agents and diverse metabolic conditions, thereby providing critical insight into gene function and pathway composition.

Acknowledgement

This work conducted by ENIGMA 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.