Multiplex characterization of microbial traits using dual barcoded genome fragment expression library in diverse bacteria

V. K. Mutalik[1] **(vkmutalik@lbl.gov),** P.S. Novichkov[1] S. Carim[2], M. Callaghan[1], H. Liu[1], M. Price[1], A.M. Deutschbauer[1], **A.P. Arkin**[1,2] and **Paul D.Adams**[1]

Lawrence Berkeley National Lab; [2]University of California at Berkeley;

http://enigma.lbl.gov

Project Goals: This ENIGMA project aims to develop a large-scale high-throughput gene-function characterization technology that is scalable, quantitative, less laborious, cheaper than currently available technologies and allow multiplex trackable quantification of microbial fitness under hundreds of conditions in different organisms. This project also aims to overcome technological limitations of RB-TnSeq[1] in characterizing essential genes and genes with no significant phenotypes in diverse conditions. One of the key goals of this project is to support ENIGMA's vision in connecting genes to function space for microbial communities and link ENIGMA environmental parameters to single-gene phenotypes in the laboratory.

ABSTRACT

Advances in DNA sequencing have had a tremendous impact on genomic research. Thousands of microbial genomes, metagenomes and single cell genomes have been sequenced, bioinformatically studied and predictions have been made about architecture of biodiversity and *in vivo* potential of microorganisms. Among this vast number of sequenced microbes, only a handful of them have been experimentally characterized for functions and fitness traits (partly because of uncultivability of most microorganisms). Current state-of-the-art technologies such as transposon mutagenesis and shotgun expression libraries coupled with next generation sequencing have been helping to fill some of the knowledge gap in sequence-to-function space. In this context, the ENIGMA SFA is developing tools and technologies for discovering and interconnecting gene functions within individual microbes and microbial communities in the field. ENIGMA is developing high-throughput tools such as randomly barcoded transposon sequencing (RB-TnSeq[1]) for cultivable microorganisms to rapidly assay loss-of-function mutant phenotypes, but it remains a challenge to infer functions of essential genes in addition to cultivating and mutagenizing individual microbial isolates from the field. One complimentary approach is to generate shotgun expression libraries from diverse microbes to perform gain-of-function assays in model organisms, but it is a very low throughput method and labor intensive. There is an urgent need for high-throughput characterization technology that is scalable, quantitative, less laborious, cheaper and allow multiplex trackable quantification of microbial fitness under hundreds of conditions in different organisms. Such a technology will provide access to the unexplored sequence-function space of hundreds to thousands of microbial

genomes including the uncultivable ones and offer novel insights into functions encoded by microbial communities.

Here we develop a next generation technology **Du**al **B**arcoded **S**hotgun **E**xpression Library Se**q**uencing (Dub-SEQ) by combining shotgun expression library method with RB-TnSeq and BarSeq techniques developed within the ENIGMA SFA. Dub-SEQ uses dual random barcoded shotgun expression libraries on broad host plasmids and increases the throughput of functional screens by using barcode sequencing (Barseq) to assess the microbial fitness. The Dub-SEQ library can have the source DNA from diverse microbial/viral genomes, metagenomes and single cell genomes and can be heterologously expressed in diverse microbes for gain-of-function screens. Each library of different size donor genomic fragments are associated with dual barcodes only once and the microbial fitness conferred by heterologous DNA is quantified by simple BarSeq protocol after assaying in different conditions expressed from different bacteria.

We have implemented the Dub-SEQ methodology in *E coli* to validate different experimental and computational methods, and to prove the interpretability and the power of the method before moving on to other library hosts and genetic material sources. Here we present pooled fitness assays across 52 stress conditions (~total 200 assays) using the *E coli* Dub-SEQ library to demonstrate the scalability of our approach. These massive selections of gene functions are of relevance to ENIGMA mission including metal tolerance/resistance, nitrate/nitrite reduction, and utilization of different carbon and nitrogen sources. As these Dub-SEQ pooled fitness assays are obtained in conditions equivalent to those carried out with RB-TnSeq[1] technology developed within ENIGMA previously, we can directly compare the gain-of-function (Dub-SEQ) dataset to the loss-of-function dataset (RB-TnSeq[1]). Here we demonstrate that the technology is scalable, quantitative, less laborious and cheaper than current gain-of-function assays.

As Dub-SEQ technology can be applied to metagenomic DNA this approach represents a high-throughput platform to link ENIGMA environmental parameters to single-gene phenotypes in the laboratory and gain insights into functions encoded by microbial communities. In addition, as the Dub-SEQ libraries can be expressed in multiple hosts, it provides a seamless platform to screen gene-functions in multiple genetic backgrounds. We believe this technology will not only support ENIGMA's vision in connecting genes to function space for microbial communities, but also broadly applicable in building functional compendium of genome fragments from diverse single amplified genomes, viruses and metagenomes from diverse environments.

References

1. Wetmore KM, et al. (2015), mBio, 6(3):e00306-15.

Funding Statement: This material 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 is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231