

The Phenotypic Landscape Of Bacteria

A.M. Deutschbauer^{1*} (AMDeutschbauer@lbl.gov), M.N. Price¹, K.M. Wetmore¹, R.J. Waters², Y. Suh¹, M. Callaghan¹, J. Ray¹, R.A. Melnyk¹, J.V. Kuehl¹, J.S. Lamson¹, Z. Esquivel¹, M.J. Blow², J. Bristow², A.P. Arkin¹, **P.D. Adams**³

¹Environmental Genomics and Systems Biology, ²Joint Genome Institute, ³Molecular Biophysics and Integrated Bioimaging,; *Lawrence Berkeley National Laboratory, Berkeley, CA, USA.*

<http://enigma.lbl.gov>

Project Goals: The Ecosystems and Networks Integrated with Genes and Molecular Assemblies (ENIGMA) program broadly seeks to understand the interactions between environmentally relevant microorganisms and their environment. One aim of this large interdisciplinary project is to rapidly bring environmental bacteria to model-organism by systematically annotating the functions of poorly characterized genes. Here, we describe a high-throughput genetic platform for generating millions of gene-phenotype measurements and how these data can be used to fill the sequence-to-function gap.

Current gene function annotation pipelines fail to identify a role for 40% genes in the typical bacterial genome. To explore the functions of these uncharacterized genes, we generated genome-wide transposon mutant populations from 25 bacteria across four bacterial classes, including 8 strains isolated from the Oak Ridge field research site. Growth of mutant populations across ~4,000 experiments, followed by random barcode transposon sequencing (RB-TnSeq), enabled over 13 million individual measurements of gene fitness, and revealed significant phenotypes for 7,375 previously uncharacterized genes. Among these genes, 4,690 had phenotypes under a specific experimental condition, or shared phenotypic patterns with a gene of known function, thus enabling specific predictions of gene function. Finally, since 11% of hypothetical genes across all sequenced genomes have at least one ortholog with a phenotype in our dataset, we demonstrate the ability of high-throughput genetics to identify roles for many of the uncharacterized proteins in bacterial genomes.

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.