Strain dynamics and functional diversity of 22 high-quality single cell genomes from ENIGMA ground water

A. Zhang[1] (anniz44@mit.edu), D. M. Needham[1,2], W. Zheng[3], S. Zhao[1], Y. Yin[1], D. A. Weitz[3], T. C. Hazen[4,5], N.S. Baliga[6], A.M. Deutschbauer[2], M.W. Fields[7], M.W.W. Adams[8], E.J. Alm[1], R. Chakraborty[2], J.-M. Chandonia[2], D.A. Elias[5], T.R. Northen[2], A. Mukhopadhyay[2], G.E. Siuzdak[9], D.A. Stahl[10], P.J. Walian[2], J. Zhou[11], A.P. Arkin[2,12] and **P.D. Adams[2,12]**

[1]Massachusetts Institute of Technology; [2]Lawrence Berkeley National Lab; [3]Harvard University; [4]University of Tennessee; [5]Oak Ridge National Lab, [6]Institute for Systems Biology; [7]Montana State University; [8]University of Georgia; [9]Scripps Research Institute; [10]University of Washington; [11]University of Oklahoma, [12]University of California at Berkeley

http://enigma.lbl.gov

Project Goals: We aim to obtain novel genomes with high quality (of completeness and contamination) from ENIGMA samples through single-cell sequencing and integrate them into KBase as good references for not-yet cultured bacteria in natural environments.

Microbial communities are diverse, dynamic ecosystems comprised of genetically diverse populations. Single-cell sequencing yields information about strain-level differences within a population that can hardly be obtained by metagenome sequencing. Thus, we aim to obtain high quality novel genomes from ENIGMA samples through single-cell sequencing and develop a pipeline to decode strain-level variations in metagenomes. Importing them into KBase as good references for ENIGMA community could have an outsize impact on environmental microbiology research.

We profiled 22 high-quality and high novelty single cell genomes (with >97% complete and < 1% contamination) from an ENIGMA groundwater well GW462 by droplet microfluidics (Microbe-seq). We observed high strain dynamics and functional diversity in these single cell genomes. To reveal the history of bacteriophage infections and adaptive immune system in ENIGMA bacterial species, we detected CRISPR loci from 4 of 22 single cell genomes (from *Aeromonas*, *Alkanindiges*, *Propionivibrio*, and *Rhodobacteraceae*) using crass crispr. Spacer sequences were extracted and searched agaisnt NCBI nt database to identify potential phages.

We identified a total of 154 unique spacers from 4 single cell genomes, indicating intense infections of bacteriophage and coevolution of bacterial host and phage. Among 154 spacers identified in single cell genomes, only 3 spacers in Aeromonas species were annotated as Aeromonas phages by NCBI. This indicates that our knowledge of bacteriophage targeting natural microbes is still quite limited. We also found the same spacer carried by multiple CRISPR.

We annotated the pathways in 22 single cell genomes by KEGG and EGGNOG databases using hmmsearch. We searched for pathways that exhibiting maximal functional diversity across single cell genomes with at least 2 fold-change and with the maximum of at least 10 genes found in a single cell. We observed high functional diversity in pathways involving antimicrobial resistance (to beta-lactam, vancomycin, and cationic antimicrobial peptide CAMP), energy metabolisms (nitrogen metabolisms, methane metabolisms, sulfur metabolisms, and oxidative phosphorylation), and xenobiotics degradation (benzoate degradation, toluene degradation, polycyclic aromatic hydrocarbon degradation, aminobenzoate degradation, nitrotoluene degradation, styrene degradation, and caprolactam degradation). This indicates distinct strategies of adaptation and diversification among bacterial species in ENIGMA environment. We further annotated nitrogen metabolisms pathways by Fama. We found a high diversity of *Burkholderiales* species in nitrogen metabolisms involving denitrification, nitrification, and dissimilatory nitrate reduction.

All single cell genomes were imported into KBase (https://narrative.kbase.us/narrative/65855) and annotated by Fama (https://narrative.kbase.us/#catalog/apps/FamaProfiling/run_FamaGenomeProfiling/), as good references for environmental microbiology research in ENIGMA community.

This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies a Science 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