

102. Novel single-cell sequencing methods link target genes with bacterial hosts

Sarah J. Spencer^{1*} (sjspe@mit.edu), Manu V. Tamminen², Sarah P. Preheim³, Zhili He⁴, Andrew Lancaster⁵, Michael Thorgersen⁵, Michael Adams⁵, Aindrila Mukhopadhyay⁶, Jizhong Zhou⁴, Eric J. Alm¹, Adam P. Arkin⁷, and Paul D. Adams⁶

¹Massachusetts Institute of Technology, Cambridge, MA; ²ETH Zürich, Switzerland; ³Johns Hopkins University, Baltimore, MD; ⁴University of Oklahoma, Norman; ⁵University of Georgia, Athens, ⁶Lawrence Berkeley National Laboratory, Berkeley, CA; ⁷University of California, Berkeley

Project Goals: We aim to deconstruct microbial ecosystems by creating assays at the scale of individual microbes. Current methods to profile single cells often restrict throughput and require expensive, technically challenging equipment. To fill this technological gap we developed a benchtop protocol that uses emulsion droplets as microreactors for targeted gene amplification. With the ability to link genomic information from single cells, a variety of applications open up, including linking functional genes to species and linking phage or plasmids to their host cells. High-throughput single-cell assays will add a missing dimension to bulk microbial surveys, which are unable to dissect flexible, dynamic microbial genomes. This will in turn inform models and efforts to observe, predict, and direct microbial function at perturbed environmental sites.

Our ability to understand and manipulate microbial ecosystems is restricted by a lack of information at scales relevant to individual microbes. This lack of information stems from the technology available; bulk sequencing techniques lose the mapping between species and their genomic content, while single-cell sequencing limits throughput. To provide an alternative, our group developed a novel method to amplify pairs of genes from single cells, while maintaining high throughput. The technique, termed Emulsion, Paired Isolation, and Concatenation PCR (epicPCR) uses emulsion droplets as nanoliter reactors to assay up to 1 million cells in parallel, with cells that can be loaded directly from environmental samples.

Our first application of single-cell gene fusions links together functional genes with phylogenetic indicator genes. We validated the protocol by assaying for sulfate reduction genes in a well- documented aquatic ecosystem. We targeted the dissimilatory sulfite reductase gene (*dsrB*) and physically linked it to a species-indicator gene (16S rRNA) within the same cell. The resulting data showed clean selection for only species carrying the *dsrB* gene, and these species all fell into well-documented sulfate-reducing clades including deltaproteobacteria. This demonstrates our ability to load a complex community, and then selectively amplify the species which carry a target gene of interest.

Expanding from our proof-of-principle, we have projects to link function to phylogeny in samples relevant to multiple campaigns. Together with Mike Adams at the University of Georgia, we are investigating metal metabolism at contaminated aquifer sites. The metabolism of heavy metals is dependent on molybdenum-utilizing enzymes, but these enzymes appear widely distributed within diverse, contaminated sampling sites. epicPCR can identify molybdenum-utilizing species within aquifers as well as within engineered bioreactors managed by the Elias Lab at Oak Ridge National Laboratory in Tennessee.

In addition, we are engineering a scale-up of this method to profile a large number of functional genes in parallel – providing a way to capture the niches of species within a community. Together with Jizhong Zhou at the University of Oklahoma, we compiled a list of prevalent metabolic genes in perturbed aquifer communities. These target genes contribute to an epicPCR panel linking functional genes to species. This panel, termed PuLSE (Panel Linking Species to Ecology), provides a detailed accounting of species along with the metabolic roles they are fulfilling within a given site.

The versatility of our emulsion-based protocol also allows for assays of spatial relationships beyond single-cell genomes. Instead of looking at how two genes from a single cell co-locate, we can assay how two genes from different cells (or plasmids, or phage) co-locate. By creating droplets containing plasmids or phage of interest, we are querying microscale host relationships. A discovery project, in collaboration with Aindrila Mukhopadhyay at Lawrence Berkeley National Laboratory, explores the host relationships of extrachromosomal plasmids. Our lab is also leading the effort to link phage functional genes (terminase, integrase, etc.) with bacterial host genes. Both projects support multiple groups from our SFA, informing studies of the groundwater microbiome as well as basic questions in evolution and community assembly.

Our ultimate goal is to offer consortium investigators a window into the myriad microscale relationships that complicate the study of complex environmental samples. We present a technique that not only removes inference from community functional analysis, but offers novel opportunities to study the spatial relationships between species and mobile genetic material.

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.