Genome-resolved metagenomics provides a comprehensive view of microbial metabolism in the subsurface

Karthik Anantharaman[1], Christopher T. Brown[2], Itai Sharon[1], Brian C. Thomas[1], Andrea Singh[1], Laura A. Hug[1], Cindy J. Castelle[1], Kenneth H. Williams[3], Eoin L. Brodie[3], Susannah G. Tringe[4], Jillian F. Banfield[1,5]

[1] Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA 94720 USA
[2] Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720 USA
[3] Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[4] Metagenome Program, DOE Joint Genome Institute, Walnut Creek, CA 94598 USA
[5] Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720 USA

Microorganisms drive transformations of carbon compounds in the terrestrial subsurface, an important reservoir of carbon on earth, and impact many other linked biogeochemical cycles. Despite this, little is known about the identity and metabolic capabilities of the vast majority of microbes in the subsurface. This lack of understanding of the microbial processes that impact carbon turnover in the subsurface has restricted the ability of climate models to capture key aspects of the carbon cycle. Here, we used a genome-resolved approach to decipher the metabolic capabilities and roles of the uncultivated microbial majority in the subsurface. We sequenced groundwater and sediment microbial communities sampled from four different geochemical environments from an aquifer adjacent to the Colorado River, near Rifle, CO, USA. Reconstruction of genomes followed by manual curation resulted in the recovery of 2,542 high-quality genomes, 18 of which are complete. Phylogenetic analyses involving 16S ribosomal RNA genes and other ribosomal proteins revealed that less than 11% of these genomes belonged to the 4 most commonly represented phyla in public databases, that constitute 93% of all currently available genomes. A large proportion of the genomes we recovered belong to phyla that were previously unknown or lacking cultured representatives. Genome-specific analyses of metabolic potential revealed the co-occurrence of important traits involving electron donors (e.g. organic carbon, sulfur, hydrogen, methane, ammonia), electron acceptors (e.g. oxygen, nitrate, sulfate), carbon fixation, and nitrogen fixation. Coupled analyses of genomic abundance and metabolic potential provide unprecedented information on microbial functional redundancy and heterogeneity across time (~6 years) and geochemical environments (sediments, groundwater). Our study significantly advances the understanding of the genetic underpinnings of microbial diversity and function in the subsurface. These predictions of microbial metabolic potential will serve as inputs into LBNL's Genome-Enabled Watershed Simulation Capability (GEWaSC) modeling effort.