

Prediction of the biogeochemical roles of uncultivated bacteria and archaea in the subsurface

JILLIAN F. BANFIELD^{1,2}, CINDY J. CASTELLE¹,
LAURA A. HUG¹, KELLY C. WRIGHTON¹, ITAI SHARON¹,
CHRISTOPHER T. BROWN¹, ROSE S. KANTOR¹,
ANDREA SINGH¹, BRIAN C. THOMAS¹, BIRGIT LUEF¹,
LUIS COMOLLI², KENNETH H. WILLIAMS², PHILIP LONG²
AND SUSAN S. HUBBARD²

¹University of California, Berkeley; jbanfield@berkeley.edu
cjcastelle@berkeley.edu, laura.hug@berkeley.edu,
wrighton.1@osu.edu, itaish@berkeley.edu,
ctb@berkeley.edu, rsgkantor@gmail.com,
andrea.singh@berkeley.edu, bcthomas@berkeley.edu

²Lawrence Berkeley National Laboratory;
birgit.luef@gmail.com, lrcomolli@lbl.gov,
KHWilliams@lbl.gov, pelong@lbl.gov,
SSHubbard@lbl.gov

Despite dramatic progress in microbiology over the past decades, vast knowledge gaps remain concerning the extent of diversity and functional capacities of organisms within the Domains Bacteria and Archaea. Also of particular note is a lack of information about the biology of subsurface environments, especially sediments, in part due to sample complexity and challenges associated with sample recovery. Further, subsurface enrichment in microbial members of candidate phyla (CP), which lack even a single cultivated representative, has prevented a comprehensive assessment of biogeochemical processes. In research carried out in an aquifer adjacent to the Colorado River in Rifle, CO, our team is combining geochemical measurements and time/depth series sampling with high-throughput metagenomic sequencing of whole community DNA, to link environmental processes with predictions of microbial metabolic potential. Importantly, even very low abundance members (below 0.1% abundance level) of highly complex communities containing thousands of species are included in this metabolic analysis. Striking findings include overall genetic novelty, prevalence of bacteria and archaea from CP and from phyla not recognized previously, and, within the CP from two domains of life, the predominance of fermentation-based metabolisms, likely symbiosis-based lifestyles, small cells, and small genomes. Data integration enables development of models that describe the flow of carbon through the communities, and can identify linkages between specific organisms and numerous biogeochemical cycles.