

Prevalence and Automated Curation of Local Errors from Metagenomic Assembled Genomes

Rohan Sachdeva^{1*} (rohansach@berkeley.edu), Livia Moura², Dylan Chivian³, **Jennifer Pett-Ridge**⁴, and **Jillian F. Banfield**^{1,3}

¹Innovative Genomics Institute, University of California, Berkeley CA; ²University of São Paulo, São Paulo, Brazil; ³Lawrence Berkeley National Laboratory, Berkeley CA; ⁴Lawrence National Livermore National Laboratory, Livermore CA

Website: <https://sc-programs.llnl.gov/soil-microbiome>

Project Goals: Microorganisms play key roles in soil carbon turnover and stabilization of persistent organic matter via their metabolic activities, cellular biochemistry, and extracellular products. Microbial residues are the primary ingredients in soil organic matter (SOM), a pool critical to Earth's soil health and climate. We hypothesize that microbial cellular-chemistry, functional potential, and ecophysiology fundamentally shape soil carbon persistence, and we are characterizing this via stable isotope probing (SIP) of genome-resolved metagenomes and viromes. We focus on soil moisture as a 'master controller' of microbial activity and mortality, since altered precipitation regimes are predicted across the temperate U.S. *Our SFA's ultimate goal is to determine how microbial soil ecophysiology, population dynamics, and microbe-mineral-organic matter interactions regulate the persistence of microbial residues under changing moisture regimes.*

Abstract: Metagenome assembled genomes (MAGs) recovered from complex communities are now the primary basis for understanding microbes in their natural environments. Consequently, the recovery of genomes that accurately reflect true biological organisms is essential. Contemporary metagenomic projects can produce thousands of genomes and the computational assemblies that comprise these mass-produced MAGs contain characteristic errors. Assembly errors perturb or even preclude functional predictions and accurate phylogenetic analyses and confound biochemical studies so limit the full potential utility of MAGs. These errors can be repaired, but because the process is time-consuming and requires human-guided manual curation, it is rarely performed. To address this, we are developing FixAME, a KBase software toolkit, for automatically curating and improving assemblies that does not require human-guided intervention. Additionally, to understand the prevalence of assembly errors, we quantified the number of assembly errors in assemblies from commonly used assembly programs across different environments. Following the complete development of FixAME we will be able to scale-up in order to curate and improve assemblies in the thousands of MAGs in public databases.

This research is based upon work supported by the LLNL 'Microbes Persist' Soil Microbiome SFA, funded by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research Genomic Science program under Award Number SCW1632 to the Lawrence Livermore National Laboratory and a subcontract to the University of California, Berkeley. Additional funds were provided by a grant to encourage new tool development for the

DOE Systems Knowledgebase, KBase, SCW1656. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344.