

KBase Science GSP: Towards a Viral Ecogenomics Toolkit at KBase

Benjamin Bolduc¹, Paramvir S. Dehal², **Jennifer Pett-Ridge³**, and **Matthew B. Sullivan^{1*}**
(mbsulli@gmail.com)

¹The Ohio State University, Columbus OH; ²Lawrence Berkeley National Laboratory, Berkeley CA; ³Lawrence Livermore National Laboratory, Livermore CA

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.*

The LLNL Soil Microbiome SFA “Microbes Persist: Systems Biology of the Soil Microbiome” seeks to understand how microbial soil ecophysiology, population dynamics, and microbe-mineral-organic matter interactions regulate the persistence of microbial residues in soil under changing moisture regimes. These processes are almost certainly shaped by viral constraints, yet we know very little about viral populations, function and activity in soils. Here we seek to establish viral ecogenomic tools at KBase to empower users to explore the role of viruses in environmental microbial communities by identifying, characterizing, and ecologically contextualizing viruses in large-scale sequence datasets. While many microbial analytical tools exist in KBase, no such viral tools were available until our project. To date, we have integrated two tools into KBase; this has required creating new data models and workflows to handle viral data. First, VIRSorter uses database matches and a probabilistic model to identify virus genomic fragments from viral and microbial metagenomes, providing users with a measure of confidence and evaluation of prophages. Second, when thousands of new viruses are discovered, there is the high likelihood that many are not yet known to science. To taxonomically classify these to approximately genus-level assignments (with confidence metrics), we have integrated vConTACT2, a gene-sharing network analysis tool, into the KBase environment. These tools, along with phage annotation tools developed at LLNL (PHANOTATE and PhATE), form a foundation for integrating the emerging viral ecogenomics toolkit into the KBase ecosystem.

This research is based upon work supported 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 Ohio State University. Work at Lawrence Livermore National Laboratory was performed under U.S. Department of Energy Contract DE-AC52-07NA27344. Work was also funded by KBase, a Genomic Science program within the U.S. Department of Energy, Office of Science, Office of

Biological and Environmental Research, funded by awards DE-AC02-05CH11231, DE-AC02-06CH11357, DE-AC05-00OR22725, and DE-AC02-98CH10886.