Viral diversity: decoding hidden potential for metabolic functions in soils

Ruonan Wu[1], Michelle R. Davison[1], Emily B. Graham[1*] (emily.graham@pnnl.gov), William C. Nelson[1], Sarah J. Fansler[1], Sheryl L. Bell[1], Iobani Godinez[1], David Paez-Espino[2], Russell Y. Neches[2], Nikos C. Kyrpides[2], Jason E. Mcdermott[1], **Kirsten S. Hofmockel[1]**, and **Janet K. Jansson[1]**

[1]Earth and Biological Sciences Directorate Pacific Northwest National Laboratory, Richland, WA
[2]Joint Genome Institute, Walnut Creek, CA

Project goals:
PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We focus on a multi-scale examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments are designed to confront spatial challenges and inter-kingdom interactions that regulate networks of biochemical reactions. The exchange among bacteria, fungi, viruses and plants are being characterized in the context of microbial metabolism and community function. These experimental data have been used to parametrize individual- and population-based models for predicting interspecies and inter-kingdom interactions. Predictions are tested in lab and field experiments to reveal individual and community microbial phenotypes. Our cross-scale experiments are coordinated together to investigate the influence of moisture on the interkingdom-interactions. Data is captured and shared through an optimized data management pipeline. Knowledge gained will provide fundamental understanding of how soil microbes interact to decompose organic carbon and enable prediction of how biochemical reaction networks shift in response to changing moisture regimes.

Although viruses are the most abundant biological entities on the planet, they have been challenging to study in soil systems. Based largely on work in other systems, soil viruses are hypothesized to be vital to ecosystem functions through transferring auxiliary metabolic genes, lysing microbial cells, and inducing microbial taxonomic turnover. Auxiliary metabolic genes (AMGs) in marine viruses, for instance, encode numerous metabolic functions involved in carbon, nitrogen, and phosphorous cycling. Yet, we have limited understanding of the identity, distribution, and ecological function of soil viruses due in part to the vast range of soil ecosystems, and until recently the lack of appropriate molecular screening tools.

Due to the lack of a universal marker gene, such as 16S rRNA that is used in prokaryote phylogeny, soil viruses have mainly been identified by screening for viral sequences from metagenomes. Recent advances in high-throughput sequencing and computational approaches make it feasible to uncover virome from the soil microbiome with high complexity. Current bioinformatic tools each have limitations on their abilities to identify viral sequences from environmental metagenomes, and viral reference databases are insufficient for soil viromes. We therefore developed the 'VirFunnel' workflow to increase confidence in soil viral assignments by leveraging multiple computational tools and viral reference databases, thereby mitigating the weaknesses of each approach with another publicly available tool. The VirFunnel workflow includes the VirSorter and VirFinder tools as well as self-curated and public-available viral

databases (JGI IMG_VR and NCBI RefseqVirus). It contains four distinct modules to (1) provide increased confidence in assignment of viral sequences from soil metagenomes (*viral mining module*), (2) assess viral diversity with a standardized and step-wise workflow (*viral clustering module*), (3) determine host composition (*host assignment module*), and (4) extract potential auxiliary metabolic genes (*AMG classification module*).

We challenged our resulting computational workflow to uncover viruses from three highly complex, deeply sequenced soil metagenomes (>1 Tbp each) that were obtained from native grassland soils in Washington, Kansas and Iowa. These soils span a range of physicochemical conditions and harbor diverse microbiomes. To date, there have been few studies of soil viruses using a metagenome screening approach, and ours represents the first for grassland soils containing incredibly high microbial diversity and complexity. By application of VirFunnel, we were able to detect 2,631 soil viral sequences representing 28% and 230% more putative viruses than solely VirSorter or Virfinder, respectively. These include many previously undescribed soil viruses.

We were particularly interested in possible ecological roles for viral clusters found in all three grassland soils. Commonly detected viral clusters contained significantly more viral sequences with more average CRISPR hits than site-specific clusters, indicating more frequent virus-host interactions. Some of these viral clusters were linked to multiple hosts across different microbial phyla, suggesting a potential for viral generalists with a wide host range in soils. Additionally, we identified potential viral AMGs previously undescribed in soils and thought to be involved in mannose degradation pathways in all soils. Auxiliary metabolic genes involved in xylan degradation were detected in WA and Kansas soils. Since xylan and mannan/mannose are common plant cell well components, multiple related AMGs detected across our three disparate grassland soils supports recent suggestions that viruses have the ability to cycle plant-derived carbon compounds in soil. Finally, we located many AMGs involved in complex carbohydrate metabolism, energy acquisition and fatty acid biosynthesis that collectively point to a central role for viruses in soil carbon cycles more broadly.

To expand our understanding of the soil virome to a global scale, we also collaborated with the JGI to collect over 3200 samples with metagenomic shotgun sequences for use in the most extensive characterization of soil viruses to date. The 3200 samples are from numerous sources including IMG, MG-RAST, EMP, NEON, and other collaborators. We plan use this dataset to define the global soil virome in terms of its size, diversity, and function. We will highlight biogeographical and eco-evolutionary patterns of viral sequences and their potential hosts, as well as identify sequences attributed to giant viruses, virophages, and AMGs. We predict that viruses are involved in multi-domain interactions in soils and play central roles in soil metaphenomes by serving as a genetic reservoir of metabolic genes.

Funding statement: *This research was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (OBER), as part of BER's Genomic Science Program (GSP), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Soil Microbiome Scientific Focus Area "Phenotypic Response of the Soil Microbiome to Environmental Perturbations." A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by OBER and located at PNNL. PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830.*