

## Untapped Viral Diversity in Global Soil Metagenomes

Jason E. McDermott<sup>1,2\*</sup> (Jason.McDermott@pnnl.gov), Emily B. Graham<sup>1</sup>, David Paez-Espino<sup>3</sup>, Colin Brislawn<sup>1</sup>, Russell Y. Neches<sup>3</sup>, Ruonan Wu<sup>1</sup>, Nikos C. Krypides<sup>3</sup>, **Kirsten S. Hofmockel<sup>1,4</sup>, and Janet Jansson<sup>1</sup>**

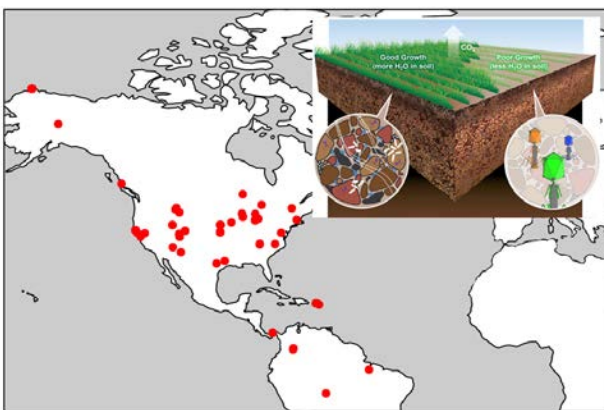
<sup>1</sup>Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA;

<sup>2</sup>Department of Molecular Microbiology and Immunology, Oregon Health & Sciences University, Portland, OR; <sup>3</sup>Joint Genome Institute, Walnut Creek, CA; and <sup>4</sup>Environmental and Molecular Sciences Laboratory, Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA

<https://www.pnnl.gov/biology/programs/>

**Project Goals: PNNL's Soil Microbiome SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture through spatially explicit examination of the molecular and ecological interactions occurring within and between members of microbial consortia. Integrated experiments were designed to confront both the scaling challenges and inter-kingdom interactions that regulate networks of biochemical reactions. Individual- and population-based models for predicting interspecies and inter-kingdom interactions were parameterized using experimental data, and predictions were tested in soil to reveal spatially explicit microbial interactions. Discoveries from controlled experiments are planned to be cross validated in the field, using moisture gradient experiments at a new local field site. Data was 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.**



Soil viruses were determined from existing soil metagenomes from >600 soil samples collected across 3 continents.

of soil viruses. Although soil viruses are particularly diverse in comparison to other habitats<sup>1</sup>, we have limited understanding of their identities and distribution. The study of soil viruses is

PNNL's soil microbiome SFA aims to understand the importance of interactions between different kingdoms (Bacteria, Archaea, Fungi and Viruses) on carbon decomposition and how changes in soil moisture, such as expected with drought, influence those interactions. Here, we specifically aimed to determine the role

particularly challenging because of the enormous range of soil ecosystems, and until recently, the lack of appropriate molecular screening tools. We aimed to overcome these challenges by screening hundreds of existing soil metagenomes that were deposited at the Joint Genome Institute (JGI) for the presence of viruses. We found that the majority of viral diversity in soil is completely novel, including new giant viruses. Additionally, we screened the viruses for auxiliary metabolic genes (AMGs) and found several AMGs that corresponded to enzymes involved in carbon decomposition. We also linked viruses to their hosts using co-occurrence networks and CRISPR spacer sequence similarity.

To identify patterns of viral biogeographic distribution we manually assigned biomes to each study using metadata deposited into Integrated Microbial Genomes with Microbiomes (IMG/M)<sup>2</sup>. Biomes considered included arctic soil, cultivated and uncultivated prairie, rhizosphere, and Canadian, Mediterranean, or tropical forests. We found that estimated viral abundances varied dramatically between these biomes, and that inferred biogeographic relationships varied as well. For example, comparisons between cultivated and native grasslands revealed that, in contrast to bacterial diversity<sup>3</sup>, diversity of viral-encoded AMGs decreased after cultivation.

We also found that most metagenomics datasets displayed a linearly increasing relationship between overall nucleotides sequenced and viral nucleotides sequenced indicating that viral diversity is significantly under-sampled. However, our study greatly advances existing knowledge of soil viral diversity and abundance and highlights the vast, uncharacterized viral diversity in soil microbiomes and their potentially important roles in carbon cycling.

## References

1. Paez-Espino, D. et al. Uncovering Earth's Virome. *Nature* 536, 425 (2016).
2. Markowitz, V. M. et al. IMG/M: a data management and analysis system for metagenomes. *Nucleic acids research* 36, D534-D538 (2007)
3. Mackelprang, R. et al. Response of the soil microbiome to cultivation in native tallgrass prairie soils of the Midwestern United States. *Frontiers in microbiology* 9, 1775 (2018).

**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.*