Title: Functional and structural characterization of soil viral auxiliary metabolic genes

Authors: Ruonan Wu[1] (ruonan.wu@pnnl.gov), John R. Cort[1], Garry W. Buchko[1], Clyde A. Smith[2], Ian K. Blaby[3], Nikos C. Kyrpides[3], Yasuo Yoshikuni[3], Janet K. Jansson[1] and **Kirsten S. Hofmockel[1]**

Institutions: [1]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99354, USA; [2]Stanford Synchrotron Radiation Lightsource, Stanford University, Menlo Park, CA 94025, USA; [3]Department of Energy, Joint Genome Institute, Berkeley, CA 94720, USA.

Website URL: https://www.pnnl.gov/projects/soil-microbiome/research

Project Goals: PNNL's Phenotypic Response of Soil Microbiomes SFA aims to achieve a systems-level understanding of the soil microbiome's phenotypic response to changing moisture. We perform multi-scale examinations of molecular and ecological interactions occurring within and between members of microbial consortia during organic carbon decomposition, using chitin as a model compound. Integrated experiments address spatial and inter-kingdom interactions among bacteria, fungi viruses and plants that regulate community functions throughout the soil profile. Data are used to parameterize 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. Knowledge gained provides 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.

Abstract: Recent research has revealed that viruses carry auxiliary metabolic genes (AMGs) that potentially contribute to soil metabolic processes while tuning the host machinery towards their own replication [1-3]. The majority of these AMGs are poorly characterized and only one, mannan endo-1,4-β-mannosidase, has been expressed and confirmed to be active in cleaving β-1,4-linked mannose, a plant-derived polysaccharide [1]. AMGs are commonly assigned potential functions based on their sequence similarity to annotated microbial genomic databases. These approaches overlook the critical motifs or catalytic sites that are key to determining if a protein is functional. Therefore, direct characterizations of AMG enzymatic functions and protein structures are of utmost importance for a mechanistic understanding of the ecological roles of soil viruses.

Here, our team focused on AMGs encoding chitosanases involved in the decomposition of chitin, the second most abundant structural polysaccharide after cellulose. We used a bioinformatics workflow that was previously developed [4] to identify a group of AMGs (glycosyl hydrolase family 75; GH75) from the largest global viral database (IMG/VR v3.0). The identified GH75 chitosanase-like AMGs were more prevalent in terrestrial environments, especially forest soils. Several chitosanase AMGs were synthesized and cloned at the Department of Energy (DOE)-Joint Genome Institute (JGI) into expression vectors. The cloned chitosanase AMGs were expressed and functionally characterized at the DOE-Environmental Molecular Sciences Laboratory

(EMSL). We were able to crystallize one of these proteins, verify its chitosanase activity and determine the first structure of any member of the GH75 chitosanase family using diffraction data collected at the Stanford Synchrotron Radiation Lightsource (SSRL). The resulting protein structure was determined at ultra-high resolution from crystals diffracting to better than 0.9 Å. The structure contained two domains: one domain containing a fold motif similar to other carbohydrate-hydrolyzing enzymes and the adjacent domain comprising a unique, previously uncharacterized fold. The active site, in a cleft between the two domains, was validated by the abolition of activity using single-site mutants of residues postulated to participate in the reaction mechanism. Further, a co-crystal structure of one of the site-directed mutants (E157Q) with chitohexaose revealed the substrate bound in the active site cleft. Because this viral chitosanase was found in forest soil and observed to be more active at an acidic pH, we speculate that viral AMG proteins may be under selection pressure to develop beneficial changes for adapting to the environment. AlphaFold2 modeling of the same GH75 chitosanase-like AMG generated a structure close to the experimental crystal structure (the root-mean-square deviation in Cα positions is 0.6 Å), demonstrating the potential of high-throughput predictions for selecting other chitosanase AMGs for functional validation. This study is the first thorough description of a soil viral AMG with detailed functional and structural characterizations. This new information will potentially enable us to better understand the metabolic contributions of soil viruses and their impact on microbiome.

References/Publications

1. Emerson, J.B., et al., *Host-linked soil viral ecology along a permafrost thaw gradient*. Nature microbiology, 2018. **3**(8): p. 870-880.
2. Trubl, G., et al., *Soil viruses are underexplored players in ecosystem carbon processing*. MSystems, 2018. **3**(5): p. e00076-18.
3. Wu, R., et al., *DNA Viral Diversity, Abundance, and Functional Potential Vary across Grassland Soils with a Range of Historical Moisture Regimes*. Mbio, 2021. **12**(6): p. e02595-21.
4. Wu, R., et al., *Moisture modulates soil reservoirs of active DNA and RNA viruses*. Communications biology, 2021. **4**(1): p. 1-11.

Funding Statement: PNNL is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RLO 1830. This program is supported by the U. S. Department of Energy, Office of Science, through the Genomic Science Program, Office of Biological and Environmental Research, under FWP 70880. A portion of this work was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by Office of Biological and Environmental Research and located at PNNL.