

Viruses may manipulate the global carbon cycle through carbohydrate active enzymes

Lindsey Solden¹, Sabina Leanti La Rosa², Simon Roux³, Cristina Howard-Varona¹, Christine Sun¹, Rebecca Daly⁴, Michael Wilkins⁴, Virginia Rich¹, Phillip Pope², **Matthew Sullivan**¹

¹The Ohio State University, OH, USA; ²Norwegian University of Life Sciences, NO; ³Lawrence Berkeley National Laboratory, CA, USA; ⁴Colorado State University, CO, USA

Project Goals: *The activity of microbes in soil profoundly affects global energy and nutrient cycles. In consequence, whatever shapes soil microbial activity shapes the world. Substantial recent work in environmental microbiology has taught us that viruses are a key driver of microbial ecology in other systems, and we expect the same to hold true in soil. But soil is a complex milieu – highly structured, chemically and physically heterogenous, and resistant to extrapolation. Thus, even as new methods have revolutionized our understanding of microbial and viral ecology in other systems, our understanding of soil microbial ecology has lagged, and our understanding of soil viral ecology is still further behind. **Our objective is to develop paradigms for understanding the role of viruses in soil ecology, and to build the tools – scalable new methods, new databases, and new model systems – to test these paradigms.***

Viruses that infect microorganisms are important players in the global carbon cycle, with relatively extensive evidence in oceans and emerging evidence in soils. Besides lysing key microbial players, viruses also metabolically reprogram their hosts – indirectly through regulatory rewiring and directly through virus-encoded Auxiliary Metabolic Genes (AMGs). In the oceans, key viral AMGs include those that impact central C metabolism, photosynthesis, and N and S cycling. In soils, however, carbohydrate active enzymes (CAZymes) may represent novel types of AMGs, though their study has only recently begun. CAZymes catalyze the breakdown, biosynthesis, or modification of carbohydrates and glycoconjugates, releasing soluble carbon to the environment, and thus representing markers of viral importance in soils.

Here, we surveyed over 1.2 million publicly available viral genomes from diverse environments to identify the frequency and ecological context of virus-encoded CAZymes. The frequency of viral-encoded CAZymes was higher than that of well-studied viral hallmark genes (e.g. terminase), with CAZymes found in ~20% of surveyed viruses. The most common (10% of the total) CAZyme domains were GH24, GH23 and GH19 which are often found in lysozyme, viral baseplate proteins and tail fibers, and likely aid the virus during host cell entry. Further, the CAZyme domains in lysozyme and baseplate proteins are specific to host cell wall type, with viruses that infect gram positive bacteria often containing GH25 and GH73, while viruses that infect gram negative bacteria containing GH23, GH24 and GH19. Using RefSeq viral isolate genomes, we developed a random forest model leveraging the frequency of CAZyme domains to predict host cell wall type. The result is a model that predicts with up to 86% accuracy, whether the virus infects eukaryotes, gram-positive bacteria, or gram-negative bacteria. Such a model can aid current prediction tools in increasing the accuracy of predicting virus-host relationships in the environment.

Other identified CAZyme domains are environment-specific, such as CAZyme domain AA10 which is detected across almost all soil environments (97% of viruses with CAZymes). This cellulose-oxidizing enzyme may aid viral hosts in degrading plant material, which is observed in other environments. Similarly, marine viral contigs contain CAZymes to break down many algal polysaccharides such as alginate, carageenan, and heparin, and viruses from both human and ruminant microbiomes contain CAZymes for degrading mucin. These CAZymes may provide the machinery for their bacterial hosts to utilize different carbon substrates, altering the carbon cycle by silently manipulating infected cells. Together, our global informatics survey of viral CAZymes helps to further understand the role of viruses in soil carbon cycling.

This study was funded by the Genomic Science Program of the United States Department of Energy Office of Biological and Environmental Research, proposal #0000248445.