EDventures Grant Final Report: Advanced Training in Genomic Analysis and the Characterization of Novel Phytoplankton Virus Genomes

Chris Schvarcz, Steward Lab, University of Hawaii at Manoa

This project was designed to bridge state-of-the-art training in genomics with a virus genome sequencing project that would further my doctoral research and contribute significantly to the field of marine virology. I proposed to sequence the genomes of 14 viruses that infect a variety of eukaryotic phytoplankton hosts and which were isolated from coastal (Kaneohe Bay, Oahu) and open ocean (Sta. ALOHA) environments. Thus far, very few phytoplankton viruses have been characterized genetically, and considering that viruses are thought to be one of the most genetically diverse communities in the ocean, there is still much we can learn and discover from sequencing new viral genomes. To develop the skills necessary to annotate and analyze these new viral genomes, I proposed to attend the Microbial Genomics and Metagenomics Workshop, held by the DOE Joint Genome Institute.

The draft genomes for the 14 virus isolates were sequenced at high coverage on the Illumina MiSeq platform. The sequencing data revealed that the viruses had diverse genome types (dsDNA, ssDNA, dsRNA, and ssRNA) and covered a wide range of genome sizes (from 2,761 bp to 1,326,596 bp), including the smallest-known as well as the two largest-known phytoplankton virus genomes. Furthermore, three of these viruses belong to families of viruses that have only previously been seen to infect plants and animals, making these isolates the first representatives of these virus families seen to infect single-celled organisms, let alone phytoplankton. Another one of the viruses lacked genes with strong homology to other known viruses and could not be assigned phylogenetically to any known virus group, thus representing a completely novel and never-before-seen type of virus.

Not only have these newly sequenced viral genomes considerably advanced our understanding of phytoplankton virus diversity, but they will also serve as reference genomes that will greatly aid cultivation-independent work. For example, nine of these viruses infect genera of phytoplankton for which there were previously no characterized viruses. Now when we detect these viruses in environmental samples, we will know specifically what phytoplankton they are infecting. Likewise, the annotation of viral genes in metagenomes will be improved, since many of these large viral genomes contain genes never seen before in viruses (in some cases more than 100 such genes).

On September 16–20, 2013, I attended the JGI Microbial Genomics and Metagenomics Workshop in Walnut Creek, CA. The workshop was an excellent learning experience, and the skills gained were used in the analysis of the genomes mentioned above. These skills will continue to aid me as I move forward with completing the genome annotation and submitting the data to the GenBank and IMG public data repositories.

In summary, this project has provided me with invaluable training in microbial genomics, and it has funded the generation of exciting novel data on marine viruses. This data has become the core of my PhD dissertation and will result in numerous publications, greatly advancing the field of marine virology.

Figure: Transmission electron micrographs of the viruses sequenced for this project; scale equals 100 nm.