CONSIDER FOR TALK

10th Annual SEA Symposium Abstract

University of South Florida

Tampa FL

Corresponding Faculty Member: Richard Pollenz (pollenz@usf.edu)



Alexis Bailey



Marian Smallin

Analysis of Conserved Genes and Comparison of Protein Sequences Across Phages From F1 and C1 Subclasses

Alexis Bailey, Marian Smallin, Joshua Heuler, Montanna Jenkins, Ariana Saldoriga, Destiny Strickland, Taylor Reid, Meghan Weber, Hannah Sheppard, Hannah Grzech, Sydni Schlosser, Daniel Nunez, Caleigh Ertel, Shreya Patel, Sydney Bass, Oliva Sciandra, Logan Hedberg, Francesca Prieto, Jeannette Myrick, James Garey, Madison Davis, Richard Pollenz

Bacteriophages are the most abundant viruses on earth that infect and replicate their DNA within bacterial and archaeal cells. The genome that is contained within each bacteriophage expresses many of the proteins that are responsible for the replication and assembly of the virus. They are a model organism for understanding genomic information, and how organisms evolve over time. Although we worked in Microbacterium foliorum during the wet lab portion of the SEA PHAGES program to provide genetic diversity, we were unable to obtain DNA that could be sequenced. Thus, we adopted TinyTim and Girr, that  are two actinobacteriphage that infect Mycobacterium smegmatis. TinyTim is a class C, subclass C1, myoviridae phage with a 1.53kb genome, and 262 predicted genes. Girr is a class F, subclass F1, siphoviridae phage with a 0.57kb genome, and 104 predicted genes. Therefore, this project endeavored to compare these genomes and identify similar and unique features. We hypothesize that genes encoding phage structural proteins will be conserved, while the majority of each genome will contain unique genes specific to the function of the phage. In addition, we hypothesize that proteins with the same function will have regions of high amino acid identity. The annotation of TinyTim revealed 223 verified coding sequences and 29 tRNAs. Girr had 101 verified coding regions and 1 tRNA. Alignment of these genomes showed minimal nucleotide minimal sequence homology. However, the annotation revealed that there are 34 genes in TinyTim that could be assigned a function compared to 32 in Girr. The common structural genes included portal, tapemeasure, major capsid, tail assembly chaperones and minor tail proteins. There also were several common enzymes including: carboxypeptidase, glycosyltransferase, helicase, and endonuclease. Alignment of the common structural proteins from TinyTim and Girr using three different programs revealed very small regions of identity and limited similarity across the entire sequence. However, secondary and tertiary structural analysis of the portal protein revealed high levels of protein structural conservation. Comparison of enzymes that had been ascribed similar functions revealed an overall higher level of sequence identity but most were still <40% conserved at the amino acid level. However, using protein alignment software, smaller regions within the proteins were identified with >60% similarity and these regions could be 3D-modeled into very similar structures even from very divergent phage classes. This research indicates the proteins with similar functions from phages in different clusters often do not have a high level of amino acid conservation and highlights the importance of rigorous analysis beyond simple amino acid comparisons when calling function. In addition, the process described in the project may assist in identifying key regions of unknown proteins for functional wet lab analysis.