CONSIDER FOR TALK

8th Annual SEA-PHAGES Symposium Abstract

Purdue University

West Lafayette IN

Corresponding Faculty Member: Kari Clase (kclase@purdue.edu)

Integration of genome sequencing and mass spectrometry to develop a pipeline for characterization of the phage bacteria system

Yi Li, Richard Chu, Matthew Holderbaum, Prasanna Janakiraman, Melissa Robin, Lindsey Saari, Gregory Bandy, Austin Cannella, Emily Coleman, Sarah Cryer, Mena Elmalh, Colleen Falk, Emma Foster, Mohammed Ghazali, Giulia Golivier, Abigail Hunnicutt, Melissa Junk, Jake Lindsey, Jenny Ly, Sean Magill, Megha Manjunath, Kelly Martin, Zachary McCrea, Suraj Mohan, Jill Osterhus, Megan Parker, Alexa Petrucciani, Zach Roberts, Srishiti Sharma, Ethan Titus, Tatumn Vernon, Yu Hong Wang, James Welch, Pei Yang, Kari L Clase

Mycobacteriophages Cosmolli16 (B1), FrenchFry (B2), Grand2040 (B1), and Hughesyang (J) were isolated during fall 2014 and fall 2015 and selected for genome annotation during spring 2016. Open reading frames (ORFs) were called using the bioinformatics software programs DNA Master, Glimmer, GeneMark, BLAST, Phamerator and Starterator. Functions of the ORFs were assigned based on homology to previously characterized proteins, location in the genome, or the presence of conserved protein motifs, using programs such as BLAST, HHPred and Phamerator. Rosebush, a well characterized phage from cluster B2, served as a reference for comparison of putative homologous proteins and functional annotation of FrenchFry.

In addition, phage proteins of phages from diverse clusters were investigated using mass spectrometry. Two software programs (X! Tandem and Mascot) with combinations of distinct databases were employed to identify peptides obtained from the phage-bacteria mixture. A comparison among the results revealed that the number of peptides identified was dependent upon the selection of software and database used to analyze and process the data.

To date, eighty four mycobacteriophages have been isolated and archived at Purdue University and nineteen have been sequenced. A pipeline for phage characterization and analysis is under current development. During the upcoming summer, genomic DNA from the remaining phages will be isolated and sequenced through the Purdue Genomics Core Facility. The core facility will use a new service, “WideSeq,” at a cost of $20.00 per sample and a requirement of 10ng of DNA. The data obtained from the sequenced phage genomes will be analyzed in conjunction with peptide data from mass spectrometry using the method above to identify novel proteins critical for mediating phage-host interactions.