DO NOT CONSIDER FOR TALK

2021 SEA Symposium Abstract

Virginia Commonwealth University

Richmond VA

Corresponding Faculty Member: Allison Johnson (aajohnson@vcu.edu)

Genome Annotation of Bacillus Phage Darren

Sameem Jaghori, Tammy Lam, Rohan Rathi, Muhammad Zain, Allison Johnson

Our SEA PHAGES course has completed gene annotation of the genome of the novel Bacillus phage Darren this spring. Genes were predicted by the tools Glimmer/Genemark, and curated by students to pick the best potential start position. We identified start positions for genes based on multiple factors: Ribosome Binding Site scores for each start position in each gene, whether the given start covered all gene coding potential as indicated by Genemark, and projected starts from Glimmer and Genemark. NCBI Blastp alignment to the best match was used to identify whether the predicted start provided a match between the 1st amino acid of the query and subject. Gene functions were predicted by the NCBI Blastp Conserved Domains tool and HHPred to define specific gene functions, and TMHMM to identify possible membrane proteins. We predicted functions based on whether conserved domains were detected by NCBI Blastp Conserved Domains Tool and used an E-value of 10^-5 or less to indicate a conserved domain hit. We also used HHPred to analyze protein functions predicted by matches the protein data bank, with scores of 90% probability and an E-value of 10^-7, and specifically chose protein functions that have a peer reviewed publication supporting the function. Finally each protein without an HHPred or Blastp Conserved Domains prediction were examined by TMHMM for transmembrane domains.  
Using Blastn, we found that Phage Darren, with a genome length of 160916 bp and a %GC of 38.7, is most closely related to Bacillus phage ALPS (Genbank accession number MN038179.1). Darren genome contains 292 predicted protein coding genes and no tRNA genes. Forty-two of those protein coding genes had functions predicted by our annotation tools.  
 We will analyze genome structure using Phamerator and compare it to other phage genomes, and share annotation highlights through our poster presentation. We are currently finalizing this annotation for Genbank submission. This will allow our findings of genomic annotations in Darren can be compared to other sequences and to share our findings with the scientific community.