DO NOT CONSIDER FOR TALK

2022 SEA Symposium Abstract

College of St. Scholastica

Duluth MN

Corresponding Faculty Member: Daniel Westholm (dwesthol@css.edu)

Comparative genomics and proteomics of actinobacteriophages Venti and LittleMunchkin

Ella K Achterkirch, Julia L Carlson, Jonah C Giermann, Ellie J Peddle, Harrison S Reinke, Rachel A Rhein, Houa Thao, Daniel E Westholm

Cluster DR Gordonia phage LittleMunchkin and cluster P1 mycobacteriophage Venti were isolated through enrichment using *Gordonia terrae* CAG3 and *Mycobacterium smegmatis* mc2155, respectively, as part of the SEA-PHAGES program. LittleMunchkin produced small clear plaques and exhibited myoviridae morphology under electron microscopy, while Venti produced small turbid plaques and exhibited siphoviridae morphology. Genome sequencing produced a 63,193 bp genome with 85 putative genes in LittleMunchkin, while Venti displayed a 46,484 bp genome with 78 putative genes. Initial comparative genomics between LittleMunchkin and other Cluster DR phages using Phamerator and Gepard dotplots revealed several regions of repeat sequences primarily in the right arm of the genome. The repeat sequences were found primarily in the genome coordinates 32,000-35,000 and 57,000-62,000. Currently, we are analyzing identifying the precise sequence of the repeats through two-way BLASTs and cross checking with promoters identified with DNA Master. Once we determine the precise sequence motifs, we will attempt to bioinformatically determine their function. In Venti, one area of interest was indicated by host-trained GeneMark output where a bimodal coding potential in ORF 43 was split by coding potential in ORF 42. This feature is present in some, but not all P1 bacteriophages. We also have further investigated the lysogeny cassette in Venti. A tyrosine integrase, repressor and antirepressor were identified during functional annotation. The att P site is localized near the integrase and is flanked by an inverted repeat, present in some but not all P1 phages. The attB site in *M. smegmatis* was localized to genome coordinates 6221028-6220990 spanning a thr-tRNA gene. Surprisingly, we have found relatively few stoperator sequences for repressor binding (7 with consensus sequence GNGCGATGTCNAG). We have also created 3d space filling models showing electrostatic potential for both repressor and antirepressor based on amino acid sequences of each. Initial analysis displays two areas of positive potential on the repressor with complementary areas of negative potential on the antirepressor, suggesting a competitive binding mechanism of antirepressor inactivation of the repressor mediated by electrostatics. Finally, we have performed LC-MS/MS analysis of infected cell pellets derived from 3 hour LittleMunchkin and Venti infections in liquid cultures. Analysis of expressed peptides will allow us to determine which genes are expressed during infection and serve as a check on the accuracy of our genome annotation.