CONSIDER FOR TALK

7th Annual SEA-PHAGES Symposium Abstract

Purdue University

West Lafayette IN

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

Isolation, Characterization and Analysis of EricMillard, a Novel Cluster J Mycobacteriophage and WaterDiva, a Novel Cluster B1 Mycobacteriophage

Jonuelle Acosta, Pedro Ajsivinac, Mark Aronson, Lauren Bailey, Ryan Benczik, Michelle Bonahoon, Yiqun Chen, Henry Clarke, Hannah Cook, Hee Gun Eom, Mokun Fatukasi, Anthony Fontana, Emma Fort, Vasupradha Girish, Nicole Griffin, Jaycey Hardenstein, Matt Holderbaum, Xueqi Huang, Nathaniel Hunnewell, Hyun Joo Kim, Arren Liu, Tamara Lofton, Kate Lowrey, Weichuan Luo, Nick Palcheff, Rushi Patel, Ben Pfister, Haley Roos, Matt Saaco, Lindsey Saari, Cameron Smith, Stephanie Tucker, Jeier Yang, Lingyu Yang, Naina Zachariah, Xinchen Zhang, Yi Li, Kari Clase

Two phages from diverse clusters were selected for annotation based upon large differences in genome size and diverse putative proteins towards the long term goal of further exploring changes in protein expression in response to changes in the lifecycle of the host bacteria, in the presence and absence of a putative canonical integrase.

EricMillard (http://phagesdb.org/phages/EricMillard/ ) and WaterDiva (http://phagesdb.org/phages/Waterdiva/ ) are novel mycobacteriophage isolated from environmental samples collected in 2012. Both phages induce lysis in the mc2 strain of M. smegmatis, have a Siphoviridae morphotype, and were sequenced using Illiumina sequencing. The genome of WaterDiva contains 68,886 base pairs of circularly permuted DNA, 103 putative proteins, GC content 66.5% and is classified as a Cluster B, Subcluster B1 with 101 total members. The genome of EricMillard contains 113536 base pairs of DNA with defined physical ends, 251 putative proteins, GC content 60.9% and is classified as a Cluster J with 18 total members.

Annotation of both phages was executed using bioinformatics programs including DNA Master, Glimmer, GeneMark, Phamerator and BLAST. ORFs were assigned a predicted function based on homology to previously characterized proteins, location in the genome, or the presence of conserved protein motifs, using programs such as BLAST, Phagesdb, HHPred and Phamerator. EricMillard contains a putative frameshift between gene 33, a tail assembly chaperone, and gene 34, the tape measure gene. Comparison of EricMillard to RedNo2 in Phamerator aided in the identification of the +1 frameshift on “slippery sequence” CCCCAAAA resulting in a change of a serine to a valine. Scanning of the EricMillard genome with tRNA machine learning algorithms ARAGORN and tRNAscan-SE, revealed a tRNA. EricMillard also encodes a putative protein with homology to the FtsK superfamily, proteins that participate in chromosome segregation, and a methyltransferase within a non-conserved region of the phage genome. Methyltransferases can methylate phage DNA and thus provide protection from host-encoded endonuclease restriction endonucleases.

Many of the putative gene products from EricMillard contained an interesting conserved region identified with HHPRED known as Outer Membrane Proteins (OMPs), specifically Beta-barrels or prepilin-type processing-associated H-X9-DG, a “targeting signal for outer membrane insertion.” This suggests that many proteins in EricMillard may be targeted to specific locations within the bacterial host cell.

Finally, several large intergenic gaps and putative overlapping genes were also identified in both WaterDiva and Eric Millard that did not share homology with similar regions of other bacteriophage in the same cluster. We are further exploring the gaps and putative reverse genes more closely by mass spectrometry and examining the expression pattern of proteins associated with the conserved OMP targeting domains.