CONSIDER FOR TALK

8th Annual SEA-PHAGES Symposium Abstract

College of St. Scholastica

Duluth MN

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



Maria R Braun



Robert S Stenseng

Evidence for revision of predicted genome annotation using tandem mass spectrometry

Ashley K Anderson, Anthony F Anselmo, Maria R Braun, Scottie E Deming, Casey J Hovland, Reed I Johnson, Courtney J Kranig, Eric W Madsen, Myranda L Marrier, Abbilyn R Primus, Teah L Rasche, Isabella Restrepo Toro, Christian W Schmidt, Robert S Stenseng, Jocelyn M Wiegand, Michael J Wolyniak\*, Daniel E Westholm

\* Hampden-Sydney College, Hampden-Sydney VA

Mycobacteriophages Lukilu, Maxxinista, and Idleandcovert were sequenced and annotated as part of the SEA-PHAGES program at The College of St. Scholastica. All phages were isolated using the enrichment technique with *Mycobacterium smegmatis* mc2155 as host. Genome sequencing and electron microscopy placed Lukilu in Cluster C1, Maxxinista in Cluster E and Idleandcovert in A3. Large turbid plaques and bioinformatically predicted gene functions associated with maintaining lysogeny supported the designation of Idleandcovert as a temperate A3 phage. The clear plaques and lack of critical lysogenic machinery (e.g. integrase) in Maxxinista and Lukilu supported a lytic classification of these phages. These properties made both phages candidates for tandem mass spec analysis of proteins expressed during infection. Actively growing *M. smegmatis* cultures were infected with Lukilu and Maxxinista at high MOI and incubated for 4 hours. The infected cells were then pelleted and then subjected to LC-MS/MS. Proteomic analysis of Lukilu and Maxxinista identified 55/225 and 36/146 of in silico predicted genes, respectively. Bioinformatically predicted start sites were verified in several genes. Interestingly, an expressed protein was detected in Maxxinista that was not called by Glimmer or Genemark. Upon inspection, this detected protein nearly completely overlapped auto-annotated gene 148. Although gene 148 had several 1:1 BLAST matches with other Cluster E mycobacteriophages in the NCBI genome database, had a strong Shine-Dalgarno score and was called by Glimmer, there was no coding potential for this gene in Genemark trained on either *M. smegmatis* or *M. tuberculosis*, and no predicted function. The tandem mass spec detected protein has a putative non-canonical start codon (not verified with sequenced peptides), moderate coding potential in Genemark trained on *M. smegmatis*, and HHpred-predicted HNH endonuclease activity. In addition, the ORF of the tandem mass spec detected protein does have BLAST matches in other mycobacterial genomes, suggesting other groups have detected it bioinformatically though auto-annotation or manual inspection of the genome. Thus, there is disagreement in the correct annotation of this region of Cluster E genomes. Our data suggests that gene 148 and its homologs are not expressed by Maxxinista or Cluster E phages, should not be reported as a gene, and instead should be replaced by the ORF whose product we detected via tandem mass spectrometry.