Cover Letter

**Annotation: Phage Shade**

November 23, 2015

To determine coding potential we used the GeneMarkS program.

Phage Shade is in the A02 cluster of phages along with recently annotated A02 cluster phages Sonny, TaeYoung & Martha. All start sites for which Shade shares homologs with these three phages are the same ones chosen for the annotated A02 phages with an amino acid alignment of 1:1 at the amino terminus except for Shade gp53 & gp68. Shade gp53 has an aa alignment of 4:2 with annotated A02 phages but start site represents the longest possible ORF. Shade gp68 as a 6:1 aa alignment with the phage Sonny homolog but the Shade gp68 start site represents the longest possible ORF.

**\*\*Could not get Starterator to work for Arthrobacter phages during Shade annotation but feel confident with our start calls since they match start sites for A02 cluster annotated genomes Sonny, TaeYoung and Martha.**

* We annotated gp8 as the Prohead protease gene since it is adjacent to the Major capsid protein gene and gives a 60% probability match to a protease domain via HHpred analysis. We note that for other annotated phage homologs in the AO2 subcluster, the Shade gp4 homolog has been annotated as the Prohead protease gene. We have annotated the Shade gp4 as containing a papain fold toxin domain.
* We annotated gp7 as a Scaffolding gene since other AO2 annotated homologs (Sonny, TaeYoung & Martha) have been designated as the gene encoding the Scaffolding protein. However, we got good hits for this gene as encoding a Mu-like Prophage I encoding protein
* We also identified Shade gp38 as an ORF with coding potential not called by Glimmer or GeneMark. This region of related annotated AO2 genomes (Sonny, TaeYoung & Martha) also have coding potential in this region upon analysis by GeneMarkS but this region in these genomes were not annotated as containing ORFs.
* Shade gp39 is annotated as a RNA polymerase sigma factor whereas the annotated A02 homologs were annotated as HTH DNA binding proteins.
* Shade gp70 we annotated as a ParB chromosome partitioning protein but the A02 annotated homologs are annotated as DNA binding proteins.

We did not detect any tRNA or tmRNA genes with web-based Aragorn. We did, however, detect a candidate tRNA coding for Gly at position 46210-46280 (Cove score = 13.0) using tRNA Scan. However, we did not annotate this region of the genome containing a tRNA gene. To do so would have resulted in an overlap with gene gp72.