**Annotation Notes: Arcadia**

Arcadia’s genome is the same size and sequence as phages Elsa and Nason except for the following: Nason has two differences 37056 (G to A) (within gp60) and 43358 (C to A) (within gp68); Elsa at positions 43271 (A to G) and 43358 (C to A).

We used GeneMarkS Output for coding potential analysis for each annotated gene.

We could not identify any tRNA or tmRNA genes using Aragorn or tRNA scan.

Genes of Special Interest

* Gp11 has a hit for a tape measure protein domain although is clearly is not the TMP. We have annotated this gene as having a TMP domain.
* We annotated gp12 as a fusion prohead protease/major capsid protein gene. The prohead protease gene is typically located close or next to the MCP. There are examples in other phages of prohead protease/MCP fusion genes. The first half of this gene gives a hit to a prohead protease and the second half of the gene gives a strong hit to an MCP.
* We called gp27 the holin gene since we did get decent hits to a holin protein.
* We did not detect any tRNA or tmRNA genes using either Aragorn or tRNA scan programs.
* We annotated gp98 as a minor tail protein. There is precedence for structural proteins located at the right hand end of phage genomes. The Cluster S mycobacteriophages have several structural proteins located near the right hand end of their genomes.
* For the vast majority of the genes except two genes the chosen start sites agreed with the recommended start sites using the Starterator program. Gp60 does not agree with Starterator but does agree with the related AM cluster Circum homolog annotated start site. Gp82 does not agree with the Starterator start site prediction but choosing the Starterator site would lose a significant amount of coding potential.
* Genes gp40, 92, 94 & 95 were not called by Glimmer or GeneMark but contained coding potential.