**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **9** |
| **Stop Coordinate** | **6339** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Gap with gene 8 39 bp** |
| **Selected Start Coordinate** | **5359** |
| **Selected Function** | **MAJOR CAPSID PROTEIN** |

**Annotation Decision #1: Is this a Gene?**

|  |  |
| --- | --- |
| **Gathering Evidence** | **Explain Your Rationale** |
| Was the gene called by an auto-annotation program (Glimmer, GeneMark)? | *Yes, both* |
| Is there evidence for coding potential? | *Both GeneMarkS and/or GeneMark-host trained coding potential maps show coding potential for gene 9.* |
| Is this gene present in other annotated genomes? | *BLAST hit Zepp\_9 e value 0*  *Phamerator (May 16th)*  *Wardwill\_9 Pham 233200*  *Zepp\_9 Pham 233200* |
| Does the gene violate any major guiding principles? | *No significant overlap with other genes. There is a gap of 39 bp with the previous gene.*  *It is long enough ( 981 bp, 326 aa)*  *Genes before and after are in the same direction.* |
| **DECISION:** | *Yes* |

**Annotation Decision #2: What is the best possible start site for this gene?**

|  |  |
| --- | --- |
| **Gathering Evidence** | **Explain Your Rationale** |
| What start sites do Glimmer and GeneMark suggest? | *Glimmer Start Coordinate: 5359*  *GeneMark Start Coordinate: 5359* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *The start site has just one possible associated Ribosome Binding site with a high score.*  *RBS 1*  *Raw SD Score -1,259*  *Z Value 3,225*  *Spacer Distance 12*  *Final Score -2,095* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *The predicted start codon is the longest ORF, the length is 981 bp. This proposed start site has a 39 bp gap with the nearest upstream gene.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *According to Starterator:*  *Candidate start sites for Wardwill\_9*  *Start 17 5359*  *The candidate start sites are conserved in other phage genomes. The start called by Glimmer and GeneMark matches the start predicted by Starterator.*  *Start 7 was called 100.0 % of time when present. It was called in 418 of the 857 non-draft genes in the pham.*  *The start number called the most often in the published annotations is 17, it was called in 418 of the 857 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *The start site is conserved in other phages genomes*  *BlastP match (best hit)*  *Zepp E value 0 Accession Number* [*QIG58098.1*](https://www.ncbi.nlm.nih.gov/protein/QIG58098.1?report=genbank&log$=prottop&blast_rank=1&RID=2RFMG4GE013)  *Alignment with Zepp*  *Q#1 MADITRADALA*  *S#1 MADITRADALA*  *The start site at 5359 is conserved in other phage genomes as indicated by BlastP.* |
| **DECISION:** | ***Keep DNA Master predicted start site 5359*** |

**Annotation Decision #3: What is the Function of the Putative Protein?**

|  |  |
| --- | --- |
| **Gathering Evidence** | **Explain Your Rationale** |
| Does this protein align with a protein having a functional assignment in BlastP (phagesDB and/or GenBank) with an alignment of 10-4 or smaller with appropriate coverage? | *The most informative BlastP match from each source*  *PhagesDB: major capsid protein from Zepp*  *E value 0*  *Q#1 MADITRADALA*  *S#1 MADITRADALA*  *NCBI: major capsid protein from Zepp*  *E value 0*  *Q#1 MADITRADALA*  *S#1 MADITRADALA*  *DNA Master: major capsid protein from Zepp*  *E value 0*  *Q#1 MADITRADALA*  *S#1 MADITRADALA*  *This protein aligns with a major capsid protein from Zepp.* |
| Does this protein align with a protein having a functional assignment in the PDB or other database in HHPred with a probability of 90% or greater with appropriate coverage? | *Selected databases: PDB\_mmCIF70\_30\_Mar, Pfam-A\_v37, UniProt-SwissProt-viral70\_3\_Nov\_2021, NCBI\_Conserved\_Domais(CD)\_v3.19*  *There is a match from HHPred*  *HIT* [*8EB4\_C*](http://www.rcsb.org/pdb/explore/explore.do?structureId=8EB4)  *NAME Major capsid protein; HK97-fold, T=9, tailed bacteriophage, VIRUS; 2.6A {Gordonia phage Ziko}*  *PROBABILITY 100*  *E VALUE 3.5e-32*  *SCORE 235.84*  *This protein aligns with a protein having a functional assignment in the databases of HHPred with a 90% probability. This protein is a major capsid protein.* |
| Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order? | *This gene is not located adjacent to genes of known functions (just a scaffolding protein upstream).This gene is in a region of the genome that shows high conservation of gene order.*  *According to Phamerator, this gene is flanked upstream by a scaffolding protein and downstream by a hypothetical protein belonging to a pham family conserved among related phages* |
| Is this gene a possible transmembrane protein? | *This gene is not a possible transmembrane protein.*  *According to DeepTMHMM there is not a transmembrane domain in this protein.*  *According to SOSUI this protein is a soluble protein.* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *YES* |
| **DECISION:** | ***major capsid protein*** |