**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **7** |
| **Stop Coordinate** | **4639** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Overlap with gene 6 1 bp** |
| **Selected Start Coordinate** | **3926** |
| **Selected Function** | **MINOR 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 7.* |
| Is this gene present in other annotated genomes? | *BLAST hit QuadZero\_6 e value e-134*  *Phamerator (May 5th)*  *Wardwill\_7 Pham 169765*  *QuadZero\_6 Pham 169765* |
| Does the gene violate any major guiding principles? | *No significant overlap with other genes. Overlap with 1 bp with previous gene*  *It is long enough (714 bp, 237 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 site do Glimmer and GeneMark suggest? | *Glimmer Start Coordinate: 3926*  *GeneMark Start Coordinate: 3926* |
| 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 (Start position 3926)*  *Raw SD Score -2,812*  *Genomic Z Value 2,495*  *Spacer distance 11*  *Final Score -3,569* |
| 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 714 bp. This proposed start site has a 1 bp overlap with the nearest upstream gene but it does not violate the Guiding Principles.* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *According to Starterator:*  *Candidate start sites for Wardwill\_7*  *Start 3 3926 bp*  *The candidate start sites are conserved in other phage genomes. The start called by Glimmer and GeneMark matches the start predicted by Starterator.*  *Start 3 was called 99.5 % of time when present. It was called in 169 of the 169 non-draft genes in the pham.*  *The start number called the most often in the published annotations is 3, it was called in 169 of the 169 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *BlastP match (best hit)*  *QuadZero E value 3e-41 Accession Number* [*QWY84452.1*](https://www.ncbi.nlm.nih.gov/protein/QWY84452.1?report=genbank&log$=prottop&blast_rank=1&RID=24YZ7XVW013)  *Alignment with QuadZero*  *Q#1 MATVPTGRM*  *S#1 MATVPTGRM*  *The start site at 3926 is conserved in other phage genomes as indicated by BlastP.* |
| **DECISION:** | ***Keep DNA Master predicted start site at 3926*** |

**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: MuF like minor capsid protein from QuadZero*  *E value e-134*  *Q#1 MATVPTGRMERE*  *S#1 MATVPTGRMERE*  *NCBI:MuF like minor capsid protein from QuadZero*  *E value 1e-170*  *Q#1 MATVPTGRMERE*  *S#1 MATVPTGRMERE*  *DNA Master: MuF like minor capsid protein from QuadZero*  *E value 0.0*  *Q#1 MATVPTGRMERE*  *S#1 MATVPTGRMERE*  *This protein aligns with a MuF like minor capsid protein from QuadZero.* |
| 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* [*Q38442*](http://www.uniprot.org/uniprot/Q38442)  *NAME GP7\_BPSPP Minor head protein GP7 OS=Bacillus phage SPP1 OX=10724 GN=7 PE=1 SV=2*  *PROBABILITY 99,84*  *E VALUE 1,2e-18*  *SCORE 144,8*  *This protein aligns with a protein having a functional assignment in the databases of HHPred with a 90% probability. This protein is a Minor head 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 downstream).This gene is in a region of the genome that shows high conservation of gene order.*  *Scaffolding protein is downstream this gene.* |
| 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:** | ***The Official Function List indicates that 'MuF-like minor capsid protein' should not be used, and recommends annotating it instead as 'Hypothetical Protein'.*** |