**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **28** |
| **Stop Coordinate** | **19117bp** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Gap 4bp** |
| **Selected Start Coordinate** | **18770bp** |
| **Selected Function** | **holin** |

**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, Glimmer and GeneMark call it gene.* |
| Is there evidence for coding potential? | *GeneMarkS and GeneMark-host show coding potencial.* |
| Is this gene present in other annotated genomes? | *PhagesDB blast hit Fulton\_26 with e value= 3e-58*  *Phamerator phage: pham 228415 (25/4/25)* |
| Does the gene violate any major guiding principles? | *Yes, the gene fails to comply with one key guiding principles* |
| **DECISION:** | *Yes, probably is a gene* |

**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: 18770*  *GeneMark Start Coordinate: 18770* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *final RBS score: -2.443*  *Z-score: 2.995*  *This is the best score.* |
| Is the predicted start codon the longest ORF? If not, does the longest ORF result in excessive gene overlap (>30bp)? | *Yes* |
| Is this start site conserved in other phage genomes as indicated by Starterator? | *Yes, is the start number called the most often in the published annotations. It was called in 72 of the 98 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master: XOE90255 (Fultom)*  *E-value: 7.5E-39*  *Q#1: MDIAFTLDPA*  *S#1: MDIAFTLDPA* |
| **DECISION:** | *Keep DNA Master pre start site 18770* |

**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? | *PhagesDB: Fulton\_26, holin*  *e-value: 3e-58*  *NCBI: holin [Microbacterium phage Fulton]*  *Sequence ID: XOE90255.1*  *e-value: 4e-72*  *DNA Master: holin [Microbacterium phage Fulton]*  *Score:345* |
| 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? | *No Quality Match*  *Name: Sodium/bile acid cotransporter; Hepatitis B virus (HBV), host entry receptor, bile acid transporter, taurocholate, Na+-c*  *Probability: 34.44*  *e-value: 350* |
| Is this gene located adjacent to genes of known function and in a region of the genome that shows high conservation of gene order? | *YES, it is between “Phame:2158 (74)” and “Phame:233367(192)” (May 16, 2025)* |
| Is this gene a possible transmembrane protein? | *Yes,*  *TMHMM:*  *# WEBSEQUENCE Length: 116*  *# WEBSEQUENCE Number of predicted TMHs: 3*  *# WEBSEQUENCE Exp number of AAs in TMHs: 64.27321*  *# WEBSEQUENCE Exp number, first 60 AAs: 41.4667*  *# WEBSEQUENCE Total prob of N-in: 0.00968*  *# WEBSEQUENCE POSSIBLE N-term signal sequence*  *SOSUI:*  *No. region transmembrane seq. type*  *1 5-27 FTLDPALVIGLLVSTVLPLLVGL Primary*  *2 37-59 IKAVLLAALALVTSLLTELGVAV Primary*  *3 66-87 DIGQGLLLAIPTFLIAVGMHYG Primary* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *Yes, “holin”* |
| **DECISION:** | *“holin”* |