**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **27** |
| **Stop Coordinate** | **18766bp** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Overlap 3bp** |
| **Selected Start Coordinate** | **18404bp** |
| **Selected Function** | ***membrane 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, 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 Quadzero\_25 with e value= 9e-64*  *Phamerator phage: pham 2158 (25/4/25)* |
| Does the gene violate any major guiding principles? | *¡No, the gene does not violate any key guiding principles* |
| **DECISION:** | *Yes, it 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:18404*  *GeneMark Start Coordinate: 18404* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *final RBS score: -3.782*  *Z-score: 2.530*  *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 40 of the 61 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master: QWY84470 (QuadZero)*  *E-value: 0*  *Q#1: MNWNDPWVPG*  *S#1: MNWNDPWVPG* |
| **DECISION:** | *Keep DNA Master pre start site 18404* |

**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: QuadZero\_25, function unknown*  *e-value: 9e-64*  *NCBI: membrane protein [Microbacterium phage QuadZero]*  *Sequence ID: QWY84470.1*  *e-value: 3e-81*  *DNA Master:membrane protein [Microbacterium phage QuadZero]*  *Score:541* |
| 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? | *Yes,*  *Name: Accessory protein Csx28; CRISPR-associated protein, ANTIVIRAL PROTEIN; 3.65A {Prevotella buccae}*  *Probability: 94.02*  *e-value: 1.6* |
| 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:235345 (80)” and “phame:235231 (117)” (May 16, 2025)* |
| Is this gene a possible transmembrane protein? | *Yes*  *TMHMM:*  *# WEBSEQUENCE Length: 121*  *# WEBSEQUENCE Number of predicted TMHs: 1*  *# WEBSEQUENCE Exp number of AAs in TMHs: 22.83424*  *# WEBSEQUENCE Exp number, first 60 AAs: 22.83307*  *# WEBSEQUENCE Total prob of N-in: 0.02150*  *# WEBSEQUENCE POSSIBLE N-term signal sequence*  *SOSUI:*  *No. region transmembrane seq. type*  *1 11-33 FVTGSLALVAAIATGIITSLFAA Primary* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *Indicate a response with a Yes or No response.*  *Once you have arrived at a functional decision, check the* [*SEA-PHAGES Official Function List*](https://seaphages.org/blog/2017/10/30/official-function-list/) *to ensure that you are following the guidelines for function naming. Functions that are not present on the approved list must be carefully vetted for approval.* |
| **DECISION:** | *membrane protein* |