**Student Gene Annotation Worksheet**

|  |  |
| --- | --- |
| **Basic Phage Information** | |
| **Phage Name** | **Wardwill** |
| **Gene #** | **26** |
| **Stop Coordinate** | **18407bp** |
| **Direction (For/Rev)** | **For** |
| **Gap (Overlap) with Previous Gene** | **Gap 63bp** |
| **Selected Start Coordinate** | **17859bp** |
| **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\_24 with e value= e-102*  *Phamerator phage: pham 228523 (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: 17859*  *GeneMark Start Coordinate: 17859* |
| Does the start site have an associated Ribosome Binding Site with a high score? | *final RBS score: -3.545*  *Z-score: 2.515*  *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 60 of the 103 non-draft genes in the pham.* |
| Is this start site conserved in other phage genomes as indicated by BlastP? | *DNA Master: QWY84469 (QuadZero)*  *E-value: 0*  *Q#1: MTKPIQVISR*  *S#1: MTKPIQVISR* |
| **DECISION:** | *Keep DNA Master pre start site 17859* |

**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\_24, function unknown*  *e-value: e-102*  *NCBI: membrane protein [Microbacterium phage QuadZero]*  *Sequence ID: QWY84469.1*  *e-value: 7e-125*  *DNA Master: membrane protein [Microbacterium phage QuadZero]*  *Score: 801* |
| 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: Green fluorescent protein, Leukocyte surface antigen CD53 chimera; CD53, tetraspanin-25, membrane protein; HET: CRO, OLC*  *Probability: 63.35*  *e-value: 120* |
| 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 “endolysin” and “2158 (74) (May 16, 2025)”* |
| Is this gene a possible transmembrane protein? | *Yes:*  *TMHMM:*  *# WEBSEQUENCE Number of predicted TMHs: 4*  *# WEBSEQUENCE Exp number of AAs in TMHs: 86.1529300000001*  *# WEBSEQUENCE Exp number, first 60 AAs: 21.63902*  *# WEBSEQUENCE Total prob of N-in: 0.98481*  *# WEBSEQUENCE POSSIBLE N-term signal sequence*  *SOSUI:*  *No. region transmembrane seq. type*  *1 39-61 YPTFDLFMVALGVCAFFIGSPLL Primary*  *2 66-88 PAWFVDSLGVSITVAAILAAVGA Primary*  *3 92-114 KLFMLEIAGKLALVFLLGGYAGT Primary*  *4 122-144 AGENGFVVLTLFSLVWLLGPRLS Secondary* |
| Is the proposed function found on the SEA-PHAGES approved function list? | *Yes, “holin”* |
| **DECISION:** | *“membrane protein”* |